Proceedings of the I·R·E

A JOURNAL of the Theory, Practice, and Applications of Electronics and Electrical Communication

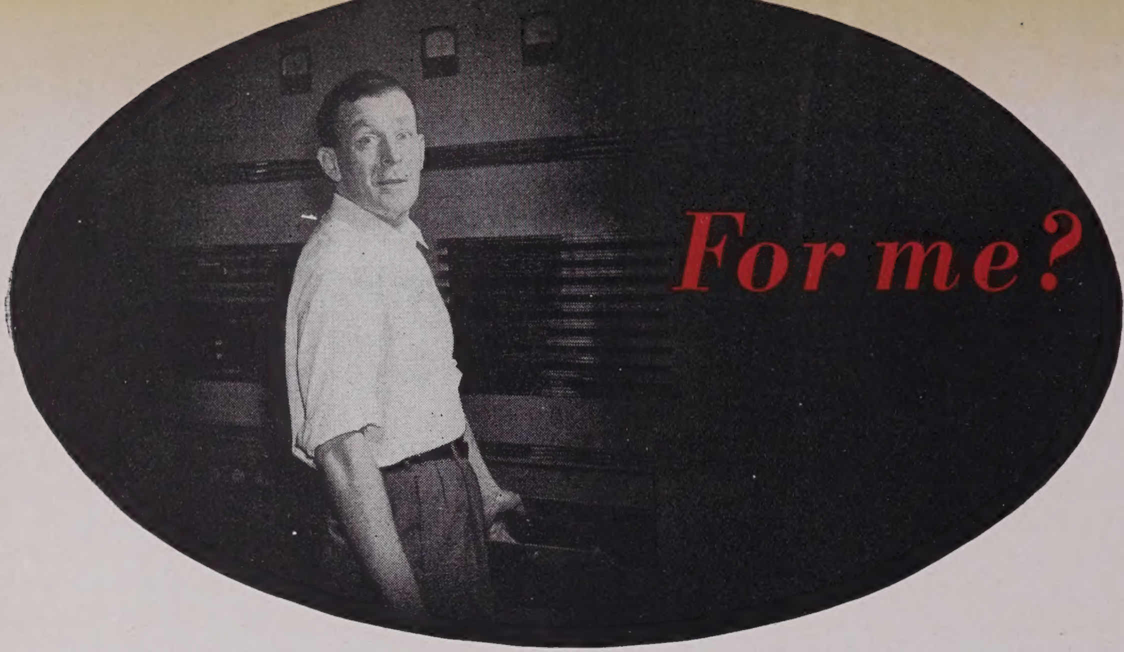
Radio Communication • Sound Broadcasting • Television • Marine and Aerial Guidance • Tubes • Radio-Frequency Measurements • Engineering Education • Electron Optics • Sound and Picture Electrical Recording and Reproduction • Power and Manufacturing Applications of Radio-and-Electronic Technique • Industrial Electronic Control and Processes • Medical Electrical Research and Applications •

WINTER TECHNICAL MEETING, NEW YORK, N.Y.
January 23, 24, 25, and 26, 1946

NOVEMBER, 1945

Volume 33 Number 11

The Institute of Radio Engineers



Yes, for you
there could very well be a citation
which would read
"For distinguished service
to the American people . . ."
. . . that is, there could be
if the nation only realized
as well as we,
who have worked with you,
what a splendid job you have done
as a radio engineer
during the emergency

If they only knew
how you overlooked the word *overtime*
and how an *eight-hour day*
lost its meaning
when we most needed
to be informed and entertained.

If they only knew
how you coddled and repaired
the irreplaceable tools
of your trade
so that not even one
valuable broadcasting moment
was lost in wartime.

If they only knew
how the station remained awake
each twenty-four hours
because of your personal effort.

. . . Well, perhaps they don't realize
to whom the thanks belong,
or their tongues don't give voice
to their feelings . . .
but in their homes and hearts
there has been mute appreciation
for the privilege you extended to all,
the privilege that could not
have been forfeited easily,
the privilege that is used so casually,
the privilege of switching on the radio.

AMPEREX ELECTRONIC CORPORATION

25 WASHINGTON STREET • • BROOKLYN 1, N. Y.

A COPY OF THIS ADVERTISEMENT, SUITABLE FOR FRAMING, WILL BE SENT WITHOUT CHARGE UPON REQUEST

Proceedings of the I·R·E

Published Monthly by

The Institute of Radio Engineers, Inc.

VOLUME 33 — *November, 1945* — NUMBER 11

Correction: The cover illustration on the October, 1945, issue of the PROCEEDINGS, entitled "Electronic Life Savers—Antipersonnel and Antitank Mine Detection" was made available through the courtesy of the Hazeltine Electronics Corporation. It is deeply regretted that mention of this co-operative action was inadvertently omitted.

Responsibility for the contents of papers published in the PROCEEDINGS rests upon the authors. Statements made in papers are not binding on the Institute or its members.

The Institute of Radio Engineers

INCORPORATED

SECTIONS

Sections	Next Meeting	Chairman	Secretary	Address of Secretary
Atlanta	November 30	R. A. Holbrook	I. M. Miles	Georgia School of Technology, Atlanta, Ga.
Baltimore	—	R. N. Harmon	F. W. Fischer	714 S. Beechfield Ave., Baltimore, Md.
Boston	November 16	C. C. Harris	A. G. Bousquet	General Radio Co., Cambridge 39, Mass.
Buenos Aires (Argentina)	—	A. DiMarco	H. Krahenbuhl	Transradio Internacional, San Martin 379, Buenos Aires, Argentina
Buffalo-Niagara	November 21	J. M. Van Baalen	H. W. Staderman	264 Loring Ave., Buffalo, N. Y.
Cedar Rapids	November 15	F. M. Davis	J. A. Green	Collin's Radio Co., 855—35 St., N.E., Cedar Rapids, Iowa
Chicago	November 16	Cullen Moore	L. E. Packard	General Radio Co., 920 S. Michigan Ave. Chicago 5, Ill.
Cincinnati	—	L. M. Clement	J. F. Jordan	The Baldwin Co., 1801 Gilbert Ave., Cincinnati 2, Ohio
Cleveland	November 27	H. B. Okeson	A. J. Kres	16911 Valleyview Ave., Cleveland 11, Ohio
Connecticut Valley	November 15	H. W. Sundius	L. A. Reilly	989 Roosevelt Ave., Springfield, Mass.
Dallas-Fort Worth	—	J. D. Mathis	B. B. Honeycutt	9025 Roanoak, Dallas 18, Texas
Dayton	November 15	L. B. Hallman	Joseph General	411 E. Bruce Ave., Dayton 5, Ohio
Detroit	November 16	L. H. Larime	R. R. Barnes	1810 Sycamore, Royal Oak, Mich.
Emporium	—	W. A. Dickinson	H. E. Ackman	West Creek, R. D. 2, Emporium, Pa.
Indianapolis	—	H. I. Metz	E. E. Alden	4225 Guilford Ave., Indianapolis, Ind.
Kansas City	—	R. N. White	Mrs. G. L. Curtis	6003 El Monte, Mission, Kan.
London (Canada)	—	B. S. Graham	C. H. Langford	Langford Radio Co., 246 Dundas St., London, Ont., Canada
Los Angeles	November 20	R. C. Moody	R. G. Denechaud	Blue Network Co., 6285 Sunset Blvd., Hollywood 28, Calif.
Montreal (Canada)	November 15	L. A. W. East	R. R. Desaulniers	Canadian Marconi Co., 2440 Trent Rd., Town of Mt. Royal, Que., Canada
New York		G. B. Hoadley	J. T. Cimorelli	RCA Manufacturing Co., 415 S. Fifth St., Harrison, N.J.
Ottawa (Canada)	November 15	W. A. Steel	L. F. Millett	33 Regent St., Ottawa, Ont., Canada
Philadelphia	December 6	D. B. Smith	P. M. Craig	Philco Corp., Philadelphia 34, Pa.
Pittsburgh	November 12	J. A. Hutcheson	C. W. Gilbert	52 Hathaway Ct., R.F.D. 1, Wilkinsburg 21, Pa.
Portland	—	Kenneth Johnson	C. W. Lund	Route 4, Box 858, Portland, Ore.
Rochester	November 15	G. R. Town	A. E. Newlon	Stromberg-Carlson Co., Rochester 3, N.Y.
St. Louis	—	B. B. Miller	N. J. Zehr	KWK, Hotel Chase, St. Louis 8, Mo.
San Diego	—	David Kalbfell	Clyde Tirrell	U. S. Navy Radio and Sound Laboratory, San Diego 52, Calif.
San Francisco	—	David Packard	William Barclay	Stanford University, Calif.
Seattle	December 13	G. L. Hoard	K. A. Moore	5102 Findley St., Seattle 8, Wash.
Toronto (Canada)	—	F. H. R. Pounsett	Alexander Bow	137 Oxford St., Guelph, Ont., Canada
Twin Cities	—	—	Merle Ludwig	Minneapolis-Honeywell Regulator Co., Minneapolis, Minn.
Washington	November 12	H. A. Burroughs	L. C. Smeby	4801 Connecticut Ave., N.W., Washington, D. C.
Williamsport	December 5	L. E. West	F. L. Burroughs	2030 Reed St., Williamsport 39, Pa.

Subsections

Subsections	Next Meeting	Chairman	Secretary	Address of Secretary
Columbus	November 9	R. C. Higgy	Warren Bauer	376 Crestview Rd., Columbus 2, Ohio
Fort Wayne	—	—	—	—
Milwaukee	—	P. B. Laeser	E. L. Cordes	3304 N. Oakland Ave., Milwaukee, Wis.
Monmouth	—	L. J. Giacoletto	C. D. Samuelson	5 Russel Ave., Ft. Monmouth, N. J.
Princeton	—	W. C. Johnson	J. G. Barry	Princeton University, Princeton, N. J.
South Bend	—	H. E. Ellithorn	J. E. Willson	1002 S. Lombardy Dr., South Bend, Ind.

The viewpoints of the Sections of the Institute are of major interest to the entire membership. Accordingly, the Chairmen of the Sections have been invited to prepare guest editorials to be published, in the form in which they are received, in the PROCEEDINGS. A broad and important aspect of professional engineering is accordingly presented in the following editorial by the Chairman of the Dayton Section of the Institute.

The Editor

The I.R.E. Philosophy

LUDLOW B. HALLMAN, JR.

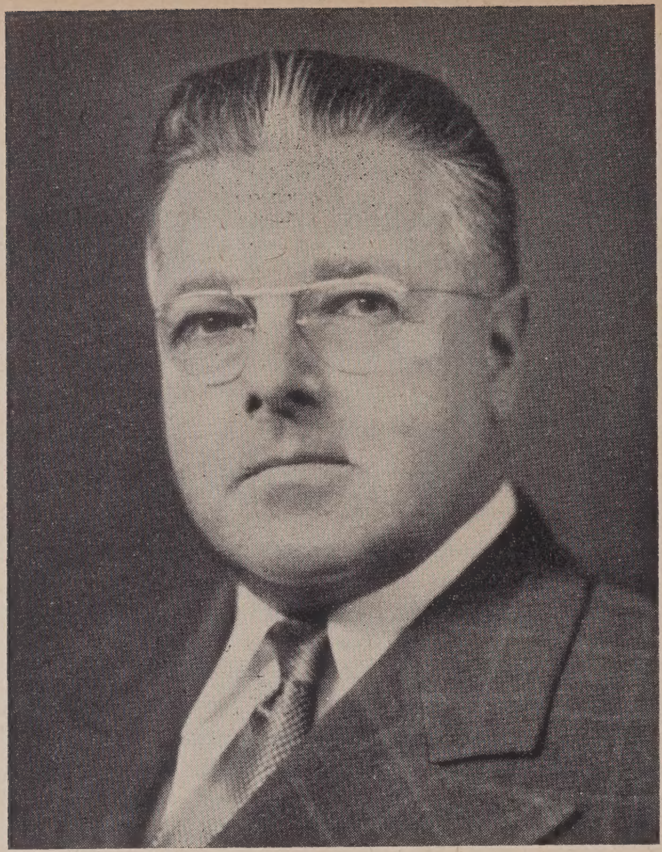
In discussing The Institute of Radio Engineers with some of my professional associates I find that many of them consider the principal function of the I.R.E. to be the publication of the PROCEEDINGS. To them membership in the I.R.E. is simply a subscription to the PROCEEDINGS.

It is not my purpose to minimize the importance of the PROCEEDINGS as a technical publication. Quite the contrary; the importance of this publication as a means of presenting and encouraging the publication of the best in electronic-and-radio engineering papers, and in advancing the art generally, cannot be overemphasized. However, it should be obvious that the basic philosophy of the I.R.E. is more profound than that of a publishing organization, since the instinctive urge which resulted in the formation of the Institute and which has made it grow and prosper is fundamental.

I suggest that the philosophy of The Institute of Radio Engineers identify itself with the basic need for developing in the radio engineering fraternity a feeling of sociological responsibility. It is a sense of this responsibility that causes the man who has made a new discovery or completed an original and important analysis to publish his findings in order that they may contribute to the over-all advancement of the art. It is, also, this same sense of responsibility that causes the radio engineer to ponder the ethics of his profession. At times it causes him consternation when he considers how the products of his imagination and ingenuity can be used just as readily to destroy civilization as to build it, dependent only on the will of the social order putting them to use.

I submit that the I.R.E. should, in addition to its other objectives, be instrumental in imbuing such a sense of sociological responsibility in radio engineers everywhere that they will demand and actively contribute to the building of a more enlightened and efficient world-wide social order; keeping pace with the advanced technological order they are so efficiently helping to fabricate.

Attending meetings and participation in the activities of local I.R.E. Sections is important in that it definitely contributes to the development of a sense of sociological responsibility among individual members of the institute and, therefore, enables the Institute as a whole to discharge more effectively its obligation in this respect. For this reason, the establishment of as many local Sections of the Institute in as many locations as can reasonably be expected to support and maintain them is to be encouraged. Furthermore, it is to be recommended that local Sections devote several meetings each year to the discussion and consideration of the sociological aspects of engineering and engineering education. Publication of papers on this general subject in the PROCEEDINGS should also be encouraged.



Lewis Mason Clement

BOARD OF DIRECTORS—1945

Lewis Mason Clement was born on January 25, 1892, in Oakland, California, and received his B.S. and E.E. degrees from the University of California in 1914.

From 1916 to 1925, he was connected with The Western Electric Company and Bell Telephone Laboratories in New York City. During this period, which included the first world war, he was engaged in the development of Government radio receivers and transmitters for the Army Air Forces and Navy, including early airplane communications systems, ground communications systems, and the CW-936 low-power combined radio, telephone, and receiver for submarine chasers. He worked on the design and supervised the installation of the Catalina Island-Los Angeles radio toll circuit in 1920, the first system to form a part of a telephone network.

In 1925 Mr. Clement became chief engineer for Fada Radio, and was responsible for the improved receiver design, test, and production methods until 1928 when he was appointed vice-president and chief engineer of the Kolster Radio Corporation. This company pioneered remote control and noise-reducing antenna systems.

During 1930 and 1931 he was with the engineering department of Westinghouse Electric and Manufacturing Company and from 1931 to 1935, he was connected with the International Telephone and Telegraph Company and its subsidiaries, The Federal Telephone and Telegraph Company in this country, and the International Standard Company in Europe. He was responsible for the design of radio receivers in Hungary, Germany, France, Italy, Belgium, and England.

Mr. Clement was vice-president in charge of research and engineering of RCA Manufacturing Company from 1935 to 1940 where he furthered the development of facsimile television, talking motion pictures, transmitters, receivers, etc.

Since 1940 to the present date, he has been vice-president in charge of research and engineering of The Crosley Corporation.

Explorations in Engineering Education*

ARTHUR B. BRONWELL†, SENIOR MEMBER, I.R.E.

Summary—**There has been a growing feeling, among engineers and educators alike, that there is a need at this time for a re-evaluation of the principles and goals of engineering education. This is due, in part, to the rapid progress of science and engineering in recent decades, necessitating a revision of our system of education along lines which will better serve society, the profession, and the engineer. In the fall of 1944, at the request of Dr. W. L. Everitt, eighteen I.R.E. Sections held scheduled meetings to discuss methods of improving engineering education. The reports submitted to Dr. Everitt's Committee on Education contained many valuable suggestions. The principal recommendations are summarized in this paper.**

Introduction

It is a self-evident principle that the social, industrial, and cultural progress of a nation is dependent, in a large measure, upon the extent and effectiveness of its educational systems. The responsibilities of leadership and the advancement of the civilization of tomorrow must inevitably rest upon the shoulders of the students of today, and their ability to cope with the problems of their generation bears a direct relationship to the quality of education and character training which they are receiving today. Leading educators have accepted this challenge and have given serious consideration to the means of improving the structure of our educational institutions as well as the educational methods.

In the areas of engineering education, the Committee on Education of the Society for the Promotion of Engineering Education has set forth clearly and concisely the broad fundamental principles which serve to chart the course of engineering education in the postwar era.[1] The course, however, is a difficult one fraught with many deceptive detours and dangerous reefs, and a chart alone, no matter how perfect, cannot guarantee a safe voyage, unless it is used in experienced hands and tempered with wisdom and discretion. The engineering and scientific professions have a realistic stake in the advancement of engineering education and, out of their combined experiences, they can contribute immeasurably to its progress.

Recently, Dr. W. L. Everitt, in an article entitled "The Phoenix,"[2] set forth a number of constructive proposals for the improvement of engineering education. He extended an invitation to the I.R.E. Sections to discuss this subject at scheduled meetings. The invitation was enthusiastically received, and eighteen Sections held meetings. The reports submitted to Dr. Everitt's I.R.E. Committee on Education bear witness to the spirited discussions that resulted. Several of the papers submitted have been published in the Proceedings. This paper presents an analysis of the principal recommendations appearing in the Sections reports, with suggestions as to how these may be incorporated in our educational system.

Fundamental Objectives

In general, there appeared to be considerable agreement as to the fundamental objectives of engineering education. These are in essential conformity with the objectives as set forth in the Society for the Promotion of Engineering Education report previously referred to, and may be summarized as follows:

1. Mastery of the fundamental physical and mathematical laws and systems of measurement.
2. Development of proficiency in the methods of engineering analysis, a comprehension of the related elements in a problem, and the ability to synthesize the various elements to obtain a practical and economical solution.
3. An understanding of the principles of organization and management, with some knowledge of production methods and costs.
4. Development of the ability of organized, logical self-expression, both written and oral, and the faculty of motivating individuals and groups of individuals.
5. Comprehension of and the ability to analyze economic and social theories and problems; an understanding of the functioning of social institutions and their influence upon society and civilization.
6. Inculcation of a philosophy of life and a set of values including a professional attitude, moral and ethical principles, a sense of responsibility, and eagerness to contribute to the advancement of society and the profession.
7. An appreciation of the higher forms of expression, including art, literature, philosophy, and music.

The first four objectives received the greatest consideration in the Section meetings since they relate more directly to the engineering development of the student. Leading educators, on the other hand, viewing world social and economic conditions with considerable apprehension, have strongly emphasized the necessity of broadening the engineering curriculum along social-humanistic lines, as set forth in the last three objectives, in order better to prepare the engineer for increased responsibilities in the management of industry and society.

Critical Selection of Instructional Staffs

Among the numerous recommendations appearing in the Sections reports, several stand out in bold relief. The necessity of obtaining competent instructional staffs

* Decimal classification: R070. Original manuscript received by the Institute, August 13, 1945.
† Northwestern University, Evanston, Illinois.
[1] "Report of committee on engineering education after the war," *Jour. Eng. Educ.*, vol. 34, pp. 589–614; May, 1944.
[2] W. L. Everitt, "The phoenix," Proc. I.R.E., vol. 32, pp. 509–513; September, 1944.

comprised of individuals who have achieved high levels of scholastic attainment, who have acquired a background of professional experience, and who have the necessary personal attributes for successful teaching, was repeatedly emphasized. It may be accepted as a foregone conclusion that the success of any educational institution is dependent first and foremost upon the competence of its instructional staff.

There are circumstances, however, which are making it difficult for engineering colleges to obtain highly qualified instructional personnel. In recent years, industry has awakened to a recognition of the need for highly trained engineers and scientists, and has drawn heavily from the reservoir of available teaching talent. Many of the better qualified college instructors have been lured into industry at salaries beyond the reach of the colleges.

The engineering colleges, themselves, have become conscious of their responsibilities in expanding the frontiers of scientific knowledge through organized research programs. This has been given added impetus by the success of war-research programs in the colleges. At present, every indication points in the direction of more widespread federal, state, and industrial subsidization of research in the engineering colleges, and many of the more progressive colleges are making plans to expand their research facilities and enlarge their research staffs. This will serve still further to deplete the reservoir of available teaching talent.

The armed-force requirements have resulted in serious reductions in undergraduate enrollment in engineering and the sciences, and the number of students taking postgraduate work has dropped to an unprecedented low. This failure to replenish the depleted reservoir of instructional talent presents a serious problem in engineering-college administration. The colleges have been compelled to recruit temporary instructors, many of whom are either not interested in or not qualified for permanent teaching positions.

In view of these considerations, it is apparent that the responsibility of the engineering colleges extends beyond the mere selection of competent instructional personnel, for it is equally important to develop high-quality postgraduate-study programs in order to enlarge the reservoir of available talent. Students who are qualified for postgraduate work should be encouraged to continue their studies toward advanced degrees, and financial aid should be made available where the need arises. A relatively large number of technician-trained men will soon be released by the armed forces. The better qualified among these should be urged to return to the colleges and continue their work toward engineering degrees.

Development of Latent Capacities

The one issue which consistently aroused widespread discussion, particularly among practicing engineers, appeared in the form of an urgent plea that more attention be given to the further development of those latent capacities of character, self-expression, professional attitude, judgment, and the ability to approach problems and situations with intelligence and self-assurance. Most engineers felt that sufficient time and emphasis, in bulk, are now devoted to the acquisition of technical knowledge and skills, but that the importance of developing the vital human attributes is vastly underrated in the engineering colleges. The opinion is frequently expressed that college instructors, in general, being academically inclined, are more concerned with imparting specific knowledge and analytical methods *per se* than in arousing a vital enthusiasm and challenging interest within the student which will serve to bring out the important human attributes.

There are those who would relegate this problem of character development to specific courses in the engineering curriculum such as psychology, public speaking, or even English, and absolve all other instructors from any responsibility in this area. Then there are others (too often the college instructors) who dismiss the problem with the terse statement that the development of these human character traits is the function of the home and the church, or that they are best developed in extracurricular activities and therefore fall outside of the realms of educational responsibilities. Such proposals, however sincerely offered, are predicated upon the false assumption that education and character training are two separate and distinct processes, each to be treated in its own sphere, and that the two are not to be combined.

In industry we find those who declare emphatically that the development of the character, abilities, and capacities of the student constitute the foremost responsibility of the colleges, and that the acquisition of knowledge and skills is of little avail unless the other attributes are developed to the fullest possible extent. Reflecting upon their own education in retrospect, they are frank to admit that most of the knowledge which they once acquired in college has long since vanished, but that the training in the powers of analysis, judgment, self-expression, and character development are the sturdy foundations upon which to build successful careers. They claim that the formal-lecture method of instruction, which seems to have become firmly entrenched in our universities, is often a boresome and inefficient method of teaching. Its principal virtue is that it contributes to the self-edification of the instructor, and serves to develop his oratorical powers. A few instructors who are endowed with exceptional personal characteristics, or who have acquired outstanding professional reputations, are able to inspire the genuine enthusiasm and vital stimulation necessary for successful teaching; but, more often than not, the converse is true.

A critic would ask us to sit in on a typical lecture (which we find to be a little on the boresome side) and observe the response of the average student. We are likely to find him listening in a half-hearted sort of way with mental processes laboring along at subnormal

efficiency, dutifully taking notes but otherwise assuming a passive interest in the whole affair. A little later they point out the same student, who has now become enthusiastically engrossed in some extracurricular or athletic activity, where he finds a vital interest and a challenge to his abilities. In contrast with the attitude of mental fatigue experienced in the classroom, we now find him eagerly planning, figuring, and working, with mental processes shifted into high gear. Our critic then asks, "Why can't we stimulate that kind of enthusiasm in the classroom?" If this were only possible, it would result in an astounding increase in educational efficiency! What are the principal motivating factors which make sports and extracurricular activities so palatable, yet which are normally lacking in the classroom? In seeking an answer to this question we discover the following:

1. In sports and extracurricular activities the student becomes an active participant and takes a challenging part in the undertaking, with complete freedom of expression and with freedom to exercise his own ingenuity. In the lecture-type classroom, the instructor takes the active lead and the student's part is subjective and passive, thus failing to arouse any vital stimulation in the student.

2. In sports and extracurricular activities we find a keen sense of teamwork and competition, which add spirit and zest to co-operative undertakings. This is usually lacking in the classroom.

3. In extracurricular activities, the student undertakes project-type responsibilities which tax his ingenuity and challenge his ability. This may appear in the homework assignments, but is not often present in the classroom.

Is it possible to devise instructional methods which will incorporate these elements and still maintain the orderly continuity and rate of progress necessary to cover satisfactorily the course material? Perhaps we can profitably take a chapter from the experience of industry, where the conference method has proven so successful in the training of foremen and executives. In a successful conference, the conference members are given advance notice of the agenda and have specific assignments to prepare. The leader opens by stating clearly and concisely the aims and objectives which are to be realized in the conference. The group then collectively formulates a method of approach and proceeds to develop the problem. Active participation is encouraged from all members of the group. The conferees present prepared reports which are discussed by the conference members. It is the responsibility of the conference leader to guide skillfully the course of the discussion with a view toward attaining the objectives.

The conference method has much to offer in our college instruction programs. It serves to stimulate interest and enthusiasm through increased participation of the individual. It serves to develop sound and logical thinking, the ability of self-expression, and a professional attitude since the individual, rather than the instructor, becomes the focal point in group discussions. This method of instruction requires careful planning and skillful guidance on the part of the instructor in order to keep the discussion moving along proper channels and to attain the desired objectives. Although the conference method is best suited to small classes, it has been successfully applied to relatively large groups on a more formal basis. The principal limitation of the conference plan lies in the fact that it is difficult to maintain the same rate of progress as in the lecture type of course, where the instructor presents the material in the orderly, logical pattern which he has found best from personal experience. However, there is scarcely a course in the engineering curricula which would not profit by a generous application of the conference method; and the returns, by way of stimulating interest and developing the potential capacities of the student, would be most gratifying.

As emphasized by Dr. Everitt, wherever possible, typical engineering problems should be assigned which combine the elements of synthesis and analysis in such a way as to tax the students' ingenuity and develop the engineering method of approach. A single project-type problem may very profitably comprise the assignment for several days or several weeks. The assignments should be such as to require frequent reference to technical publications, and the students should prepare material for oral presentation to the group. The broader aspects of economic, social, and ethical considerations should be freely discussed wherever they arise. Engineering and science should be taught as a dynamic body of knowledge which is constantly expanding, and the limits of present-day knowledge and direction of expansion should be discussed.

In the project type of laboratory, we find a means of acquiring experience in the engineering method and developing creative thinking. Here, a small group of students selects a project-type experiment in consultation with the instructor, plans the method of approach and equipment to be used, and carries it through to a successful conclusion. The number of experiments conducted in this type of laboratory is considerably smaller than in the cookbook variety of experiment, where the student is handed all of the ingredients and a set of instructions on how to mix them to produce the desired results. However, the mental processes which the student goes through in developing the project-type experiment are similar to those experienced in engineering practice.

Universities are frequently open to the criticism that they fail to train adequately new instructors in efficient and effective teaching methods. Too often it is taken for granted that a young instructor who has a master's degree or a doctorate degree has enough intelligence and resourcefulness to devise successful teaching techniques of his own. The new instructor, however, is likely to find the demands upon his time weighing heavily. He is

expected to continue his advanced studies, carry on research work, and develop himself professionally, in addition to his routine departmental responsibilities. Consequently, he finds little time to develop the fine art of teaching. Successful teaching is a skilled art, and one in which the new instructor can profit immeasurably by the experience of others. In many cases, the efficiency of instruction could be greatly improved through definitely scheduled and planned conferences in which the new instructors would have an opportunity to discuss educational methods, as well as their own particular problems, with experienced instructors.

Correlation of Mathematics, Physics, and Engineering Courses[3]

The need for better correlation of the mathematics, physics, and engineering courses is always a favored topic in discussions on engineering education. Here we find two camps with diametrically opposed viewpoints. The first camp contains critics who contend that mathematics courses are all too often taught in the nebulous abstract by puritanical mathematicians who are scrupulously careful not to contaminate a beautiful science with practical applications. The critics argue that the student fails to appreciate a beautiful science in its pure form, and merely learns to perform mechanical manipulative processes without the slightest comprehension of the ultimate use. Later, the student encounters engineering applications, and is deeply distressed and confused by the fact that the mathematics here bears little resemblance to that through which he had previously struggled. To the critics in this camp, the solution is quite obvious. They contend that mathematics should be taught as an applied science, with each new mathematical principle introduced by way of a physical problem. It is contended that the physical concepts serve as a visual aid to guide the student over the difficult mathematical steps toward an eventual solution.

In the second camp we find the equally ardent opponents of the applied mathematics viewpoint. They argue that the student who receives the mathematical principle and the physical application simultaneously is compelled to grasp two concepts simultaneously, neither one of which is familiar to him. This mental juxtaposition often results in confusion and the failure to recognize the fundamental mathematical principle. Furthermore, they contend that the applications are invariably special cases of a general theory, and that mathematics taught in this way is deprived of its generality, rigor, and clear-cut demonstration of fundamental principles.

A compromise approach might offer a reasonably satisfactory solution to the problem. Thus, the fundamental mathematical principle could first be presented in its pure form. This would then be followed by illustrative examples in which the principle is applied to physical problems well within the grasp of the student. Such an approach would require that the mathematician know something about the applications of mathematical principles. Typical illustrative examples could readily be furnished by the physics and engineering departments, and it might prove beneficial to have the mathematics instructor sit in on physics and engineering courses.

There is also a need for closer correlation among various engineering courses. In the fields of statics and dynamics, fluid mechanics, thermodynamics, and electrical engineering, we find many similarities in analytical methods. The courses should be taught in such a manner as to emphasize these similarities, and thus break down the compartmentalization barriers which have grown up between our engineering courses. Advanced courses in engineering mathematics have already accomplished much in serving to break down these departmental barriers.

Degree of Specialization

The issue which invariably strikes the sharpest discords of opinion in round-table discussions among engineers and educators is the question of whether the major emphasis in the engineering curricula should be placed upon (1) a broad general engineering education; (2) highly specialized engineering courses of the practical variety; or (3) highly analytical but fundamental engineering education.

An executive engineer in a large corporation states that his experience has shown that the majority of graduate engineers eventually become engaged in design, production, or sales work, where they have little use for the highly analytical course material taught in many engineering courses. On the other hand, most engineers need more knowledge of materials, mechanical design, production methods, costs, and management. These are the elements with which the average engineer works in everyday engineering practice. A prominent radio engineer arises to decry the fact that graduate electrical engineers who enter the radio profession may know Maxwell's equations forward and backward, but they flounder in hopeless confusion when confronted with a job of designing simple circuits for a radio receiver. He contends that there should be an electronic curriculum for students desiring to enter the radio profession, and that this curriculum should emphasize the design of radio and electronic circuits, materials, and equipment.

A leading educator hastens to remind us that, up to the last couple of decades, the majority of foremost engineers in this country, the Steinmetzes and the Timoshenkos, received their education in European universities, where the major emphasis was placed upon analytical subject material. He points out that, in recent years, American engineering colleges have placed increasing emphasis upon analytical methods, with the result that

[3] E. A. Guillemin, "Co-ordination of the work of the physics, mathematics, and electrical engineering staffs," *Jour. Eng. Educ.*, vol. 35, pp. 401–406; March, 1945.

the American trained engineer now attains a professional stature comparable to the European trained engineer. Therefore, the continuation of American engineering supremacy requires that greater stress be placed upon fundamental analytical material in the engineering curriculum with a unifying of the underlying continuity throughout all of the physical sciences and engineering courses. Furthermore, he contends that the practical aspects of engineering can be learned much more thoroughly and with considerably greater facility in industry; whereas, in college, the student should concentrate upon developing analytical abilities.

He would remind us that much of the material which we now consider as basic in the engineering curriculum was either unknown twenty years ago or was frowned upon as being nebulous, unintelligible theory which had no place in the engineering curriculum. As an example, he cites the theory of modulation which was first presented by Carson scarcely twenty-five years ago. At that time, modulation was considered a mystifying and highly theoretical subject which only the physicists and mathematicians could hope to fathom. Yet, today, this theory is given in practically every undergraduate radio-engineering textbook, and even the die-hard radio engineer would insist upon its inclusion in the electrical-engineering curriculum. In view of the spectacular progress being made today in physics and engineering, we can set it down as a foregone conclusion that the engineering curriculum twenty years hence will contain many new fundamental scientific and engineering principles, and it is the responsibility of the engineering instructor to ferret out the truly fundamental concepts and put them up in presentable form for the undergraduate curricula. He declares emphatically that, when the time comes when we cease to bring down into our undergraduate curricula new scientific and engineering fundamentals, then engineering progress will have reached stagnation and we will be well on the road to decadence.

A distinguished individual arises. "Gentlemen," he begins, "the evidence presented in this round-table discussion clearly and decisively points to one inevitable conclusion. In order that our engineering curriculum encompass the rapidly expanding technological advances and, at the same time, provide the breadth necessary for a proper understanding of social, economic, and political elements in our society of today, the engineering curriculum must be extended to include at least five or six years. The training of doctors and lawyers extends over a period of seven to ten years. Is it, therefore, asking too much of the engineer to devote, let us say, six years of his life to his professional development?"

A conservative engineer calmly counters with the reminder that there are other factors which must be considered when making decisions affecting the duration of the undergraduate engineering curriculum. Thus, the majority of engineering students receive their baccalaureate degree between the ages of 22 to 25 years. This is the age of life at which young men should seriously think of getting married, raising families, and embarking upon a professional career. Regardless of how thoroughly the engineer is trained in college, he still must spend an apprenticeship in industry before he can become an experienced and valued engineer. It is better that he complete this apprenticeship at an early age. Doctors and lawyers start their principal life work at about the age of thirty. This is entirely too late a start for the average engineer, both from a sociological and a professional viewpoint.

He would also like to remind us that those of us who expect the young student, emerging from a chrysalis of college education in a somewhat faltering state of mind, to be thoroughly grounded in scientific and engineering principles; to have a broad comprehensive training in design, materials, production processes, and costs; to have an intensive training in a field of specialization; to have an adequate background of social and humanistic studies; and to possess professional maturity and experienced judgment are expecting far too much of human nature. He points out that the problem of inculcating this conglomeration of virtues in an orderly and coherent pattern in one stable individual is quite analogous to undertaking the intricacies of teaching muscular co-ordination to a paralyzed centipede! Even the mature engineer seldom possesses all of these multitudinous virtues. He expresses the firm belief that the engineering curriculum should concentrate largely upon the logical and orderly development of scientific and engineering fundamentals, with a moderate amount of emphasis placed upon the broad cultural subjects and business courses. The engineer can then acquire specialization and a further knowledge of business methods during his apprenticeship in industry. Our present system of daytime and evening graduate courses offers ample opportunity for the engineer to expand his education further along scientific, business, or cultural lines.

And so the battle royal rages on!

At this point, we turn to the engineering educators and ask of them what they propose to do in this dilemma. In the S.P.E.E. report, we find clear-cut recommendations for many of our inquiries. This report recognizes the need for dividing engineering students into three categories, classified as (1) the largest portion comprising students pursuing a normal engineering curriculum; (2) students preparing for operation and management pursuits; and (3) the scientific-technical group. All three groups would receive the same broad general courses in the physical sciences and engineering fundamentals. In the senior year the second group would take additional courses in production methods, management, cost accounting, and other business courses. The third group would pursue a planned and co-ordinated sequence of science and engineering courses extending through the last two years of the undergraduate curriculum and into the graduate program.

This committee recommends that the undergraduate

curriculum include a thorough grounding in the broad fundamentals of the sciences and engineering, an integrated sequence of courses in social, humanistic, and business studies, and a moderate amount of specialization along the lines of further developing fundamental principles in a particular field of specialization. It recommends that some of the more specialized engineering material now taught in the undergraduate curricula be transferred to the graduate program, in order to clear the way in the undergraduate curricula for higher priority material of a more fundamental nature. It is fully cognizant of the fact that this program will not turn out experienced and mature engineers, and industry must therefore assume the burden of training the student along specialized lines required for the particular industry.

Rather than extend the duration of the undergraduate engineering curricula,[4] the committee recommends concentrating attention upon (1) making a critical selection of course material; (2) providing a better correlation of material and improving the underlying continuity of courses in the engineering curriculum; and (3) improving the effectiveness and efficiency of instruction methods.

The committee has strongly emphasized the importance of providing for a carefully planned and co-ordinated sequence of social-humanistic courses. It is becoming increasingly apparent that many of the larger and more difficult problems facing the engineering profession today are of a social, economic, or political nature. The engineer must take his full share of responsibility in the solution of these problems, or democracy and the engineering profession will surely suffer. Education for democracy is fully as important as education for a profession. This portion of the educational program would include a fundamental treatment of (1) the individual and his environment, including an analysis of factors contributing to the rise and degeneracy of ancient and modern civilizations; (2) the functioning of social and industrial institutions, including labor and management problems; (3) economics of our modern industrial society; (4) an analysis of political systems; and (5) moral, ethical, and social philosophies.

The issues discussed here are but a few of the more controversial problems facing the engineering colleges. Space limitations preclude a discussion of many of the problems considered at the Sections meetings. Several of the less controversial issues included the necessity of improving the selection of students admitted to engineering schools; the need for better mathematical preparation in the high schools; the desirability of having students and faculty members alike supplement their engineering education with experience in industry; and the need for stimulating research in engineering colleges as a desirable adjunct to engineering education.

The problems discussed here are, in general, not amenable to clear-cut decisions based upon tangible evidence alone, but rather, must necessarily reflect the combined experiences of industry, educators, and the engineering profession. A free interchange of ideas is vitally necessary in order to assure an enlightened viewpoint as a basis for intelligent decisions. In the final analysis, however, the destinies of engineering education lie in the hands of the individual instructors in the colleges; theirs is the privilege and the obligation of assuring American supremacy in engineering education.

[4] For a discussion on extension of undergraduate program, see H. J. Gilkey, "Discussion of SPEE committee report on engineering education after the war," *Jour. Eng. Educ.*, vol. 35, pp. 332–334; January, 1945.

Radar in the United States Army*

History and Early Development at the Signal Corps Laboratories, Fort Monmouth, N. J.

ROGER B. COLTON†

Summary—**The evolution of radar technique is traced and the radar-development program of the United States Army Signal Corps at the Signal Corps Laboratories, Fort Monmouth, New Jersey, described from its inception to America's entry in the War. Two radars developed by the Signal Corps Laboratories during this period, SCR-268 and SCR-270, are described in detail.**

Introduction

RADAR, the chief electronic weapon of the war, has a much longer history than is realized. The scientific concepts on which radar is based go back to the last century, and the military concept of its use arose at least 15 years ago. So generally was the idea appreciated that most of the major belligerents had radar equipment ready before their entry into the war. The United States was no exception. Both Army and Navy had radar ready. On December 7, 1941, the Army had 580 sets on hand. An Army radar, the SCR-270, gave warning of the impending attack on Pearl Harbor.

The military and naval accomplishments of radar, before and since that time, are of the first magnitude. Time and time again, winning a victory has proved easier because we have had more and better radar than

* Decimal classification: R561×R090. Original manuscript received by the Institute, September 18, 1945. Presented, New York Section, September 5, 1945.
† Headquarters, Army Air Forces, Washington 25, D. C.

our enemies. In critical stages of the war, notably the battle of Britain and the early naval engagements in the Pacific, radar has proved to be one of the decisive factors second only to guns, armor, ships, planes, and the men who fought the battles.

The value of radar has been no secret to our enemies. Throughout the war there has existed a technical race to achieve and maintain radar superiority. While this race continued, many of the interesting achievements could not be described. But sufficient time has passed to permit disclosure of early work. It is my purpose in this paper to describe that portion of this early work with which I am most familiar, namely the radar-development program of the United States Army Signal Corps.

The Evolution of Radar Technique

Radar's primary purpose in war is to give knowledge of enemy activity. It does so by exploring the region of battle with a directed beam of radio energy, and by detecting the echoes which arise when the beam en-

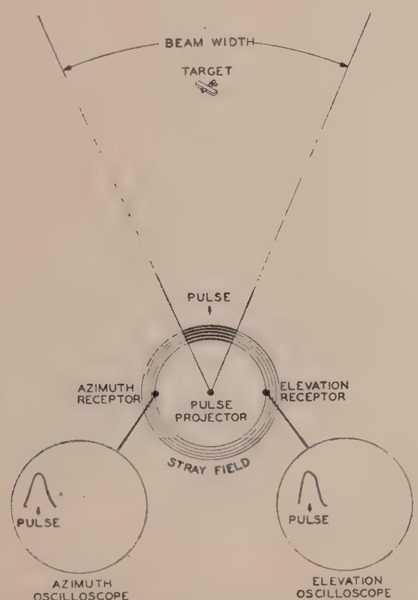


Fig. 1—Diagram of pulse in transit, showing record on oscilloscope. Taken from Signal Corps Laboratories 1937 annual report.

counters an enemy target. To detect targets at great distances, which is necessary to give adequate warning of their approach, it is necessary to transmit at the highest possible power and to receive the echoes with the most sensitive possible receiver.

Radar is, in this respect, one of the most inefficient devices known to electrical science. Radar transmitters customarily have peak power output in the tens or hundreds of kilowatts, and the effectiveness of this power is increased several hundred times by directive antennas. But the power received back from a target, at the maximum range at which the target is detectable, is measured in micromicrowatts, or roughly a millionth of a millionth of a millionth of the power transmitted.

It is no wonder, therefore, that radar development has demanded the most advanced techniques known to radio engineers and scientists. But the basic idea is simple. It depends on five requirements: (1) the production of a high-power beam of radio energy, capable of being moved about in search of targets; (2) the transmission of short bursts or pulses of energy, with comparatively long quiet periods between them during which the echoes may be detected; (3) the reflection of these pulses by the target; (4) perception of the echoes by a highly sensitive receiver and cathode-ray indicator; (5) meas-

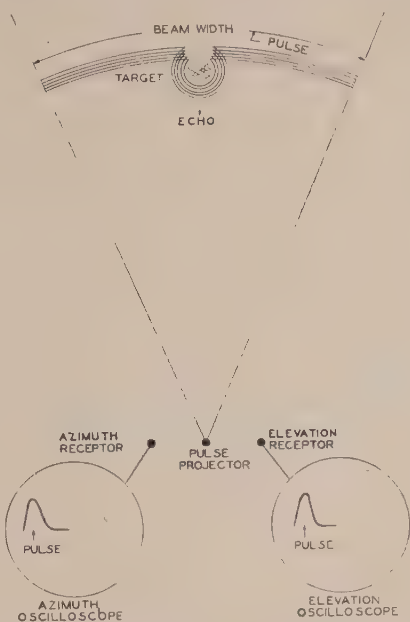


Fig. 2—Diagram of pulse in transit, showing reflection at target. Taken from Signal Corps Laboratories 1937 annual report.

urement of the time between transmission of a pulse and reception of the echo, to determine thereby the distance to the target.

Fig. 1 illustrates the first step in sending out a pulse of energy for the purpose of detecting the target airplane. In Fig. 2 the pulse of energy has reached the target airplane and is being reflected in all directions from the airplane. Fig. 3 shows the reflected energy arriving back at the position of the radar equipment.

All of these requirements could be met, in some degree, very early in the history of radio science. Hertz demonstrated in 1885, using 66-centimeter radio waves, that beams could be formed and that solid objects would reflect them. Moreover, when the identity between light and radio waves was established, it became clear that a radio wave, reflected back on itself, would create a wave-interference pattern, and that this pattern would in itself be evidence of the reflecting object.

This wave-interference detection method, the forerunner of the pulse method, was reported by various groups of workers in widely different applications, in the early 1920's, both in the United States and abroad.

In the latter 1920's the pulse method of detection of the ionosphere was introduced by scientists in this country.

The Signal Corps program leading directly to the development of radar began in 1931 with the transfer of "project 88" from the Office of the Chief of Ordnance to the Signal Corps Laboratories at Fort Monmouth. This project was entitled "Position Finding by Means of Light," the word "light" being interpreted in the broad sense to include infrared and heat rays. Later, the project was extended to include very short radio waves. At the start, the activity of this project was confined to the development of infrared devices, to detect the heat of aircraft engines and the funnels of surface ships. These devices were, in fact, included as a part of the first radar equipment built at the Laboratories.

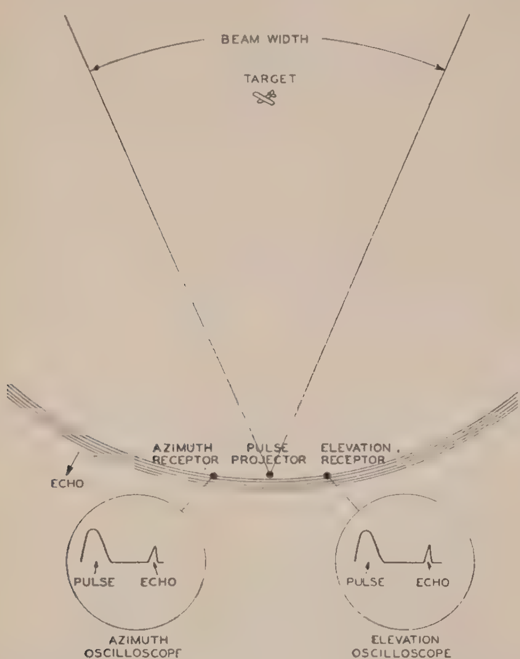


Fig. 3—Diagram of arrival of echo at receptor. Taken from Signal Corps Laboratories 1937 annual report.

In 1932 it had become clear that infrared radiation suffers from obstruction by clouds and that infrared receivers do not have sufficient sensitivity to provide detection at great distances. Consequently, in 1932 and 1933, the Signal Corps Laboratories undertook a systematic survey of the production of very short radio waves, and subprojects were set up to study "radio-optical detection and position finding." The information provided by other agencies was studied and had considerable influence on subsequent Signal Corps activity.

The first experiments were conducted in 1933 with continuous-wave equipment, employing a 9-centimeter magnetron of Westinghouse design. Ranges of a few hundred yards were obtained on moving vehicles.

In 1934, experiments were made with somewhat similar Radio Corporation of America equipment (Figs. 4 and 5). Both of the above equipments were of too low power for practical results.

The first proposal to use pulses within the Signal Corps organization was made in July, 1934, in the annual report of the Signal Corps Laboratories, as follows: "It appears that a new approach to the problem is essential. Consideration is now being given to the scheme for projecting an interrupted sequence of trains of oscillations against the target and attempting to detect the echoes during the interstices between the projections. No apparatus for this purpose has yet been built."

Up to this time the Signal Corps work was in the hands of Major Clayton and Messrs. Anderson, Zahl, Golay, Hirschberger, and Noyes. These gentlemen laid a good foundation and continued to assist in the program.

Fig. 4—Early experiments by Signal Corps Laboratories on 9-centimeter continuous-wave equipment. Twin Lights, New Jersey August, 1934.

Fig. 5—Early experiments by Signal Corps Laboratories on 9-centimeter continuous-wave equipment located on Signal Corps boat. August, 1934.

In 1936, funds in the amount of $80,000 were made available by the War Department for active prosecution of airplane-detection work during the fiscal year 1937. Before the apparatus was built, an important decision was made. It was decided to abandon the attempts to use microwaves, because transmitter power and receiver sensitivity were inadequate, and to use frequencies in the 100-megacycle region, for which negative-grid tubes of high power output were available.

At this time, responsibility was transferred by Lieutenant Colonel Blair, the Laboratory Director at that time, to Major Corput and Mr. Watson, who remained in charge thereafter and who deserve the major credit for the successful conclusion of the project.

A breadboard model was constructed early in 1936 on 133 megacycles, and later the frequency was changed to

110 megacycles. This construction marked the beginning of the Signal Corps development of the SCR-268 and SCR-270, the first United States Army radars.

The 1936, equipment had a power of 75 watts, and was pulsed at a rate of 20,000 pulses per second. It comprised, in addition to the transmitter: a phasing unit, keying unit, superregenerative receiver, cathode-ray indicator, and simple directive antennas.

The early pulse equipment was unsuccessful at first because the receiver used was incapable of recovering its sensitivity immediately after being blocked by the transmitted signal. Using the continuous-wave method, beat notes were detected between direct and reflected waves. The transmitter and receiver in this case were separated by several miles. This equipment was successful in November, 1936, in detecting aircraft if they were close to the line connecting transmitter and receiver, but its inability to indicate the direction of the aircraft was a serious stumbling block, so the method was dropped and work on the pulse method continued. The superregenerative-receiver recovery time was shortened by increasing the quench frequency. Superheterodyne receivers were also developed with low-Q circuits for the same purpose. In December, 1936, using these receivers, aircraft were successfully tracked over ranges up to seven miles using the pulse method. A yagi transmitting antenna was used to provide directivity and the separation between transmitter and receiver was about one mile.

Mr. Hessel designed the superregenerative receiver which as usual was the forerunner of the superheterodyne (designed by Mr. Moore). Perhaps here a little credit should be given to Major E. H. Armstrong, for the basic design of receivers of this kind.

This first success of the pulse equipment was marred by the inaccurate indication of direction, so attention was directed toward improved antennas. Arrays of half-wave dipoles were constructed early in 1937 to provide a high degree of directivity in azimuth (horizontal angle with respect to north). The array consisted of 12 dipoles, each $4\frac{1}{2}$ feet long, arranged in two horizontal rows of six dipoles each. When two such arrays were used, one on the transmitter and one on the receiver, a B10-B bomber was tracked up to 23 miles, the error in angle being of the order of 7 to 8 degrees. This was a great improvement over previous results, but still not of sufficient accuracy for the intended purpose. Bell Laboratories gave us help in this task, and throughout all our later development.

The next step was to build three different arrays, one for the transmitter (5 dipoles high by 2 dipoles wide), a receiving array for azimuth indication (4 dipoles high by 8 dipoles wide), and a second receiving array for elevation indication (5 dipoles high and 2 dipoles wide). The large size of these arrays made it impracticable to mount them all on a single structure, so three mounts were built, each capable of directing the associated array in any direction. The three mounts were connected by selsyn indicators so that all three could be pointed in the same direction at once. Meanwhile, a new transmitter of 5- to 10-kilowatts peak power was constructed and two superheterodyne receivers provided, one for each receiving array. The transmitter was pulsed at a rate of about 8000 pulses per second. During this period we received considerable help from the Radio Corporation of America.

Many successful tests were conducted with this equipment in the early months of 1937, culminating with a demonstration to Mr. Harry Woodring, the Secretary of War, on May 26. On this occasion the direction of the receiving arrays was transmitted directly to a searchlight. As the radar arrays were tracked on the target (a B10-B aircraft) by the operators, the searchlight followed the aircraft and could illuminate it at the will of the officer in charge—always provided the radar was doing its work. When Mr. Woodring attended, the equipment worked exceptionally well, to the surprise and relief of all concerned. Planes were detected and tracked at distances as far as 20,000 yards (11 miles). Actually using an RCA transmitter, which was not used in the demonstration, we once covered a distance of 32 miles during our preliminary tests.

Fig. 6—First 110-megacycle test of complete radar system together with heat detector. This is the first radar system in which azimuth, elevation, and range were obtained from one equipment. Signal Corps Laboratories, May, 1937.

Fig. 6 shows the complete layout of the equipment used in the May, 1937, demonstration. On the left is the elevation receiving antenna. The second antenna is the azimuth receiving antenna. The next object in the foreground, which looks like a searchlight, is the heat detection unit; and the next antenna, which, however, does not show very clearly in the picture, is the transmitting antenna.

Fig. 7 is a close-up view from the back side of the azimuth receiving equipment. The receiver is in the little box below the axle of the mount. The soldier sitting with his back to the camera is looking into the oscilloscope by means of which he sets the antenna in azimuth. The box on the right houses the Rangertone range unit.

The transmitter assembly shown in Fig. 8 was developed by Mr. Marks. Fig. 9 is a close-up view of the elevation receiving equipment. The receiving equipment here was designed by Mr. Moore.

In the foreground of Fig. 10 is the range unit developed by Rangertone, while Fig. 11 shows Zahl's heat

FIG. 7—Detailed view of azimuth antenna used in system shown in Fig. 6. May, 1937, tests.

detector, developed by Zahl and Golay. Fig. 12 shows the organization of the personnel and equipment.

The directivity of the arrays still remained unsatisfactory. In the demonstration just described, a thermal

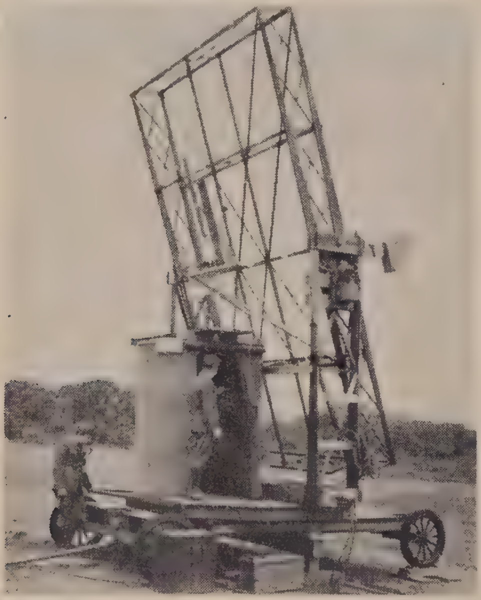


Fig. 8—Detailed view of transmitting antenna. May, 1937, tests.

detector was used to improve the sharpness of angular indication. The susceptibility of the thermal indicator to interruption by weather difficulties required that some means be found of improving the angular precision of the radar. The means was found in the technique known as lobe-switching.

The difficulty was that a sufficiently narrow beam could not be produced, at a frequency of 110 megacycles, by an array of practical size. The direction of the target was obtained by moving the beam past the target and attempting to determine the direction at which the received signal was strongest. Since the beam was broad (20 to 30 degrees at the half-power points) the direction of maximum signal was correspondingly imprecise, generally not better than 10 degrees. The military requirement was for angular precision of 1 degree or better.

The clue to the answer was found in the "A–N" radio-range beacon, used in navigating aircraft, developed by Lieutenant-Colonel Murphy at the Signal Corps Aircraft

Fig. 9—Detailed view of elevation antenna. May, 1937, tests.

Radio Laboratory. In this system, a precise path is found by overlapping two broad beams and following a path in the overlap region where the two beams have equal strength. The general principle is illustrated in Fig. 13.

If the direction of the target is outside the overlap region (along the line *OGFE* for example) there is a wide disparity between the signal strengths of the two beams, whereas equal signal strength is found only along the line *ODC*, i.e., the center of the overlap region.

The first application of this principle to the Signal Corps radar equipment took the form of two arrays,

Fig. 10—Detailed view of range-finding equipment located in rear of azimuth antenna. May, 1937, tests.

mounted at an angle to each other. The input of the receiver was switched rapidly from one array to the other, and the two signals were displayed side by side on a cathode-ray tube whose sweep circuit was displaced synchronously with the switching between arrays. The two arrays were then moved, as a unit, until the two signals had the same amplitude. The overlap region of the two arrays then pointed directly at the target, within an error which was very small compared with the width of the beams employed. This elementary type of lobe-switching was introduced in August, 1937, in the form of a twin array consisting of two units, each four dipoles wide, the two units diverging in the direction by 20 degrees.

The provision of the double-lobe pattern in both the vertical and horizontal planes from a single antenna was contemplated by the Signal Corps group from the inception of the lobe-switching idea; however, there were many problems other than the consolidation of an-

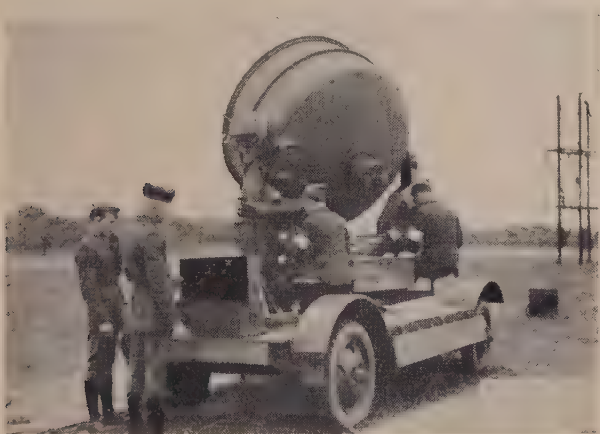


Fig. 11—Detailed view of radar and thermal-control unit. May, 1937, tests.

tennas to be solved, and it was not until 1938 that a systematic attack was made on this problem. The method finally adopted was based on the simple theory that if an antenna is fed from the right-hand side the reaction will be different from, but symmetrical to, that which will occur when it is fed from the left-hand side. This reasoning eliminated consideration of all complicated formulas for matching, phasing, and the like. Upon trial of this method it was immediately found that good lobe-switching resulted, but there still remained many problems in connection with keeping lobes steady in relation to the axis of the antenna as the antenna was rotated either in elevation or azimuth. Several methods were found for doing this to a high degree. As finally resolved, the problem turned out to be constructing the array in such a way as to eliminate, as far as possible, all but one type of radiation. In the actual case of the SCR-268, this meant the elimination, as far as possible, of all radiation other than horizontal radiation. The lobe-switch and indicating design work was done by Messrs. Moore, Cole, Deisinger, and Slattery.

While this development was under way, other branches of the Army were kept advised of progress, and specific military uses for the new equipment were considered. Two such uses were evident: (1) detection and tracking of aircraft as an aid to searchlight control and antiaircraft fire; and (2) warning of the approach

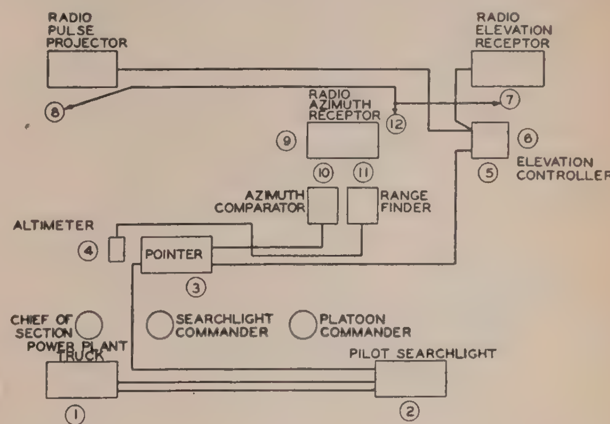


Fig. 12—Block diagram showing arrangement of equipment used in the Signal Corps tests of May, 1937. Word "pointer" refers to thermal-detector.

of enemy aircraft at great distance. The first application was fulfilled by the development of the SCR-268 series of radars, and the second by the SCR-270 series.

Officially, the development of the SCR-268 began in February, 1936, when the Chief of Coast Artillery prepared a set of "military characteristics" describing the desired aircraft detector. The requirements specified were: (1) ability to operate in daylight or darkness; (2) ranges up to 10,000 yards in mist, rain, fog, smoke, or 20,000 yards under average atmospheric conditions; (3) the position of the aircraft to be determined to an angular accuracy of 1 degree in azimuth and elevation; and (4) the distance to be accurate to one per cent of the range. Either thermal detectors or radio detectors or a

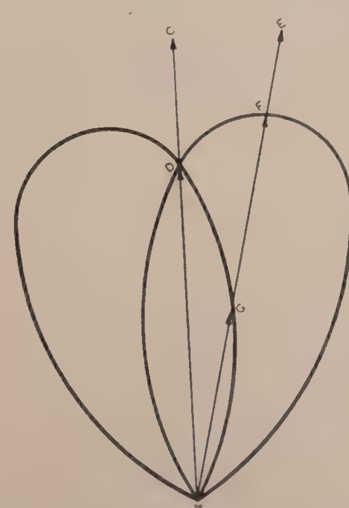


Fig. 13—Extract from an early Signal Corps report showing the principle of direction finding with two antennas displaced with respect to each other.

combination of the two was acceptable. In its attempts to solve the problem posed by the Coast Artillery, the Signal Corps Laboratories designed and built three models of the SCR-268 radar. The last of these met the specifications and was put into large-scale production.

The first service-test model was the SCR-268-T1. This radar employed the equipment described in the preceding paragraphs. The frequency was 110 megacycles; the transmitting, azimuth, and elevation arrays were separated, and each receiving array was of the twin type providing double-lobe tracking. The equipment included a heat-detection device to track the aircraft by the heat of its engines. This thermal element was mounted separately and resembled a searchlight in appearance. The equipment was ready for demonstrations early in 1938, and Coast Artillery personnel were trained in its use. In November, 1938, the equipment, being mobile, was moved to Fort Monroe, Virginia, to the Coast Artillery Board, to see whether it met the requirements of military use. During two weeks of continuous testing, tracking B-10 and O-25 aircraft, the following facts were apparent: The radar had a range of 40,000 yards, twice that set up in the military characteristics. The angular errors as reported by the Coast Artillery Board averaged about 4 degrees in azimuth and $2\frac{1}{2}$ degrees in elevation, as against the 1 degree specified in the characteristics. The average error in measuring the distance was of the order of 700 yards. The equipment was judged useful in all respects except the angular indications. During these tests the thermal element of the radar system proved to be extremely accurate, but since its field of view was limited and since it was estimated that, because of clouds being interposed between the airplane and the radar, the thermal indicator would be useless about 75 per cent of the time, its development was put on low priority; and although not actually abandoned as a research problem, the idea of its use in connection with radar detection of airplanes was given up. During the tests, a B-10B aircraft participating in the tests encountered a wind of 120 miles per hour and was blown out to sea without the pilot's knowledge. The radar discovered this fact. After the identity of the plane was checked by requesting it to circle, the pilot was directed back to the coast. This was the first instance of radar navigation in the United States Army. Also during these tests, it was suggested to the Coast Artillery Board that an attempt be made to detect the bursting of antiaircraft shell. The radar was successful in indicating the burst of 3-inch antiaircraft shells at a range of several thousand yards.

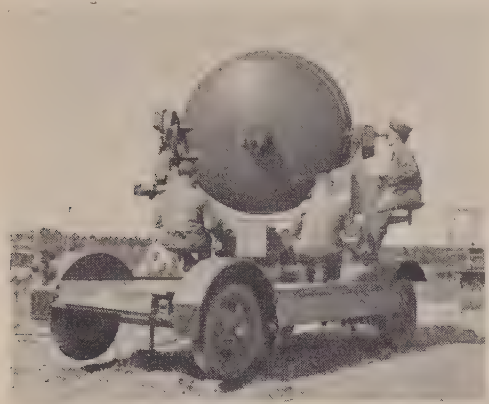


Fig. 14—First service test of the SCR-268-T1. Conducted at Fort Monroe, Virginia, October, 1938. Figure shows the radar control station built around the heat detector which was used for accurate positioning in the experimental model.

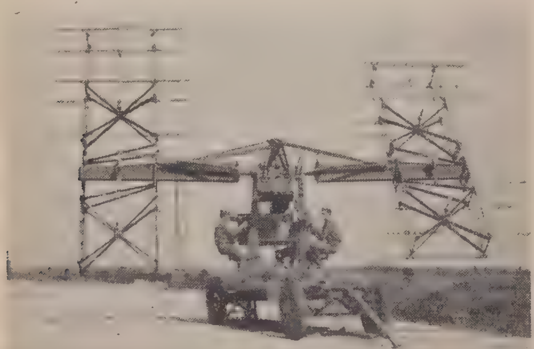


Fig. 15—Elevation antenna SCR-268-T1. October, 1938.

Fig. 16—Transmitting antenna SCR-268-T1. October, 1938.

Fig. 17—Azimuth antenna SCR-268-T1. October, 1938.

Figs. 14 to 17, inclusive, show all the units of an SCR-268-T1 radar. Fig. 15 shows an elevation receiving equipment of the SCR-268-T1. In Fig. 16 is shown the transmitting equipment of the SCR-268-T1, and Fig. 17 shows the azimuth receiving equipment of the SCR-268-T1, while the transmitting equipment of the SCR-268-T1 is shown in Fig. 18.

Meanwhile work was under way on the second model, SCR-268-T2, similar in general form but operating at 205 megacycles. The separate mounts for transmitter, azimuth, and elevation arrays were retained. As early as 1937, work was under way at a frequency of 240 megacycles, the purpose being to permit reduction in the size of the arrays. Aircraft echoes were demonstrated to the Secretary of War with 240-megacycle equipment designed by Messrs. Hessel and Slattery, in May of that year.

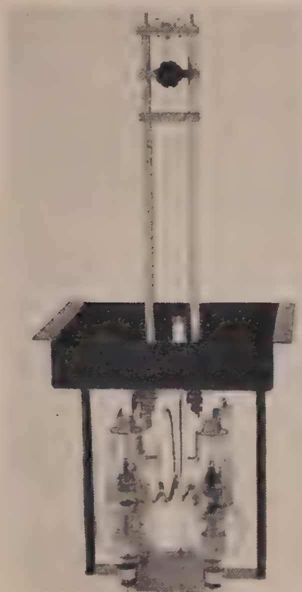


Fig. 18—Transmitter made up of four 806 vacuum tubes, tubes, SCR-268-T1. October, 1938.

An equipment on 240 megacycles, used in May, 1937, is shown in Fig. 19. It had a range of about 6000 yards, and represents the first work in the 200-megacycle band at the Signal Corps Laboratories.

One problem in this development was that of achieving sufficient transmitter power to meet the minimum range requirements. By September, 1939, a 205-megacycle model (SCR-268-T2) similar to the 110-megacycle SCR-268-T1 was ready for service test but was abandoned in favor of the SCR-268-T3, which had all the arrays on a single mount. In addition, a much more powerful transmitter was constructed for the 268-T3 through the use of a "ring" circuit incorporating eight tubes, operating at 205 megacycles and designed by Mr. Baller. The transmitter tubes were Eimac 100TL tubes redesigned. The number of tubes was later increased to 16. The T3 model was completed and service tested by April, 1940, found acceptable in all respects, and standardized for production. In August, 1940, a contract was let to manufacture the T3 model. In December, 1940, the first of 18 preproduction models was completed by the Signal Corps Laboratories, under the supervision of Mr. Vansant, prior to the delivery of production in January.

The first trainload of production models of 268's on

Fig. 19—240-megacycle radar equipment tested in May, 1937.

the way to troops is shown in Fig. 20. By February, 1941, 14 commercial models were delivered and shipped out to Army units. Since then a total of 2974 SCR-268 radar units has been produced. Production engineering and manufacture was by Bell Laboratories and Western Electric. Production was terminated in March, 1944, with the advent of superior equipment.

In July, 1941, seven SCR-268 radars arrived in Panama and were set up for tactical use by October of that year. Two sets arrived in Iceland in August, 1941, with the troops sent there to protect the North Atlantic sea route, and by December, 1941, 16 sets were in use by Coast Artillery troops in Hawaii. Since then the SCR-268 has served in all theaters of War, not only for search-

Fig. 20—First rail shipment of 13 radar sets SCR-268 leaving Fort Hancock for tactical units.

light control but also as a gun-laying set. Inevitably, in view of its wide deployment, SCR-268 was captured, and by 1944 the Japanese had paid us the compliment of copying it.

Detailed Description of SCR-268 Production Model

A view of a production model of the SCR-268 is shown in Fig. 21.

The radar equipment is carried on a trailer on which is mounted a rotatable pedestal. The pedestal carries the three antenna arrays, the transmitter, receivers, and indicators. Reading from left to right are the azimuth receiving array, transmitting array, and elevation receiving array. Behind each receiving array is the corresponding receiver. Near the center are three cathode-ray indicators, with seats in front of each for the operators. The handwheel in front of the elevation operator

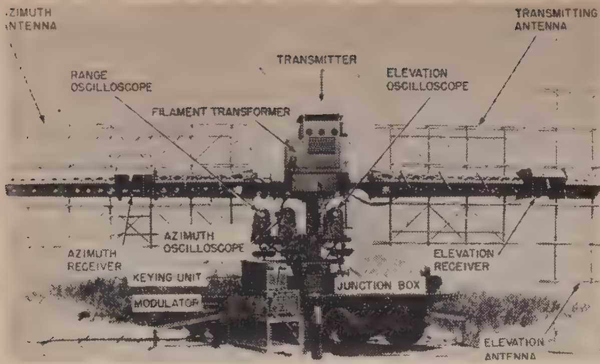


Fig. 21—Broadside view of the SCR-268 radio.

Fig. 22—Oscilloscope operators on an SCR-268 radar maintain watch near Menella, Italy. Nearest the camera is the range scope, next, the azimuth scope, and, at the right, the elevation scope.

Fig. 23—Radio set SCR-268 with 5-man crew in operation on an Italian hillside. The three operators seated on the mount see indications of the airplane echo on cathode-ray oscilloscopes. One operator tracks the aircraft in azimuth (direction in degrees from north); another operator tracks in elevation (angular height); and the third measures the range.

permits him to raise or lower the arrays in elevation. By turning these controls, the operators keep the arrays, including the transmitter array, pointed at the target. The third operator measures the range of the target by turning a range-unit handwheel which displaces the echo on his oscilloscope to the hairline and also transmits the range to the altitude converter. The mount was designed by the Breeze Corporation. In addition to the apparatus shown, a separate trailer carried a Le Roy gasoline-engine generator, a rectifier designed by Bell Laboratories, and frequency-measuring equipment designed by Fred M. Link. Mr. Slattery had charge of system design for the SCR-268-T3.

Fig. 22 is a close-up view of the range oscilloscope, range handwheel, and range operator. Figs. 23 to 27 are views of the SCR-268 in the field.

The original range unit built for the SCR-268 early in 1937 used a resistor-capacitor type of phase shifter,

Fig. 24—Radio set SCR-268 in action at Nettuno, Italy, assisting antiaircraft guns to shoot down German night bombers over the Anzio beachhead in 1944.

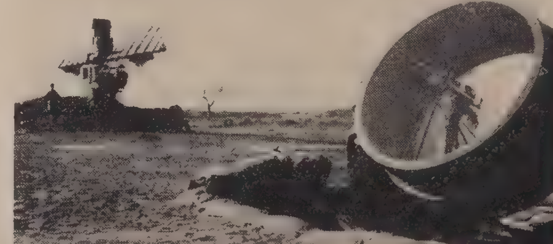


Fig. 25—In North Africa, radio set SCR-268 served to control searchlights like that in foreground, so that when the light is turned on it will flash instantly on an enemy airplane. These sets were used extensively in safeguarding our forces in North Africa.

designed by Rangertone. Later, a Helmholz inductor-capacitor type of phase shifter designed by Dr. Anderson was used.

In operation, the azimuth operator turns his handwheel back and forth, causing the arrays to scan from left to right and back over the sector in which enemy planes may be expected. When he detects an echo, he causes the azimuth array to bear directly on the target by equalizing the double-lobe signals. Thereupon the elevation and range operators go to work, the elevation operator adjusting his wheel until his double-lobe signals are equalized, and the range operator determining the range. The values of range, azimuth, and elevation

thereby determined are fed through selsyn drives to the associated searchlight. In the case of antiaircraft-gun control, the target co-ordinates are fed to a director which introduces the necessary "lead" in advance of the target position to allow for the speed of the aircraft and the time of flight of the shells.

In addition to the pedestal trailer, a power-supply trailer was furnished. In later modifications, four trucks were used to supply power and transport the gear. In this case the whole equipment, including trucks and spare parts, weighs about 20 tons, and its power is furnished by a 15-kilovolt-ampere gasoline-engine generator. The weight and bulk of the equipment, while very great compared with more recent sets, have not prohibited rapid installations of the SCR-268. It has been set up, many times, within several hours of being put ashore on a beach.

The transmitter proper, located at the top of the pedestal, generates radio-frequency pulses at a peak power of approximately 50 kilowatts. The transmitter consists of 16 triodes in a ring circuit, the plates and grid of adjacent tubes being connected through half-wave transmission-line tuned circuits. This type of circuit avoids putting the tube capacitances in parallel and permits the high power to be achieved at the frequency of 205 megacycles. The 16 tubes are necessary to obtain sufficient emission to produce the 50-kilowatt pulses.

The pulses themselves are about 5 microseconds long and occur at a rate of about 4100 pulses per second, that is, one pulse every 240 microseconds. The transmitter is on the air, therefore, only about 2 per cent of the time. During the remaining 98 per cent, the receivers are active and listening for echoes. The listening interval determines the maximum range of the set, since the signal must reach the target and return to the receiver between transmitted pulses. Since radio waves travel at a rate of about 330 yards per microsecond, in 240 microseconds the signal travels a total of about 80,000 yards; i.e., 40,000 yards to the target and 40,000 yards return trip. If targets are observed at a greater distance than 40,000 yards (23 miles), the echo is obscured by the next transmitted signal or it may arrive during the next succeeding listening interval. In this latter case, the signal may be tracked, but the range measurement is in error by 40,000 yards, a fact which is usually evident from the weak condition of the signal.

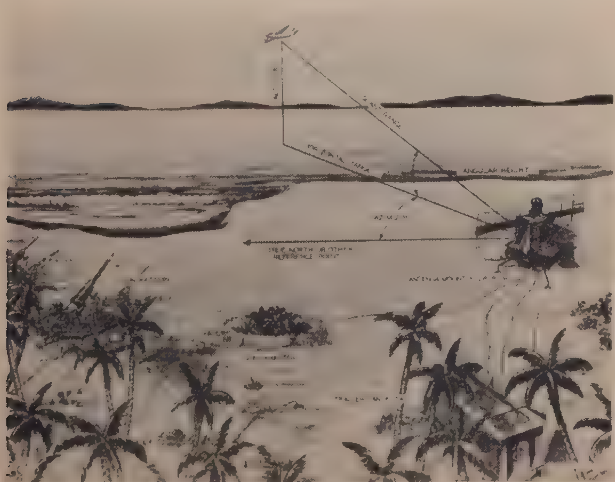


Fig. 26—Radar control of antiaircraft artillery fire is illustrated in this sketch of the SCR-268 radar supplying firing data to a gun director. The electrical impulses from the radar are fed to the gun director, which automatically points the guns and sets the shell fuses for the correct altitude.

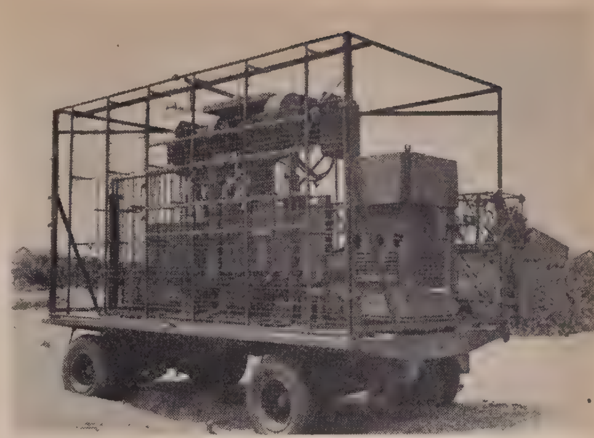


Fig. 27—Late model of radio set SCR-268 trailer packed for transport, less canvas cover.

In a similar manner, the length of the pulse, 5 microseconds, determines the minimum range at which targets can be detected, since no echoes can be seen while the transmitter continues to transmit. In addition, the recovery time of the receivers is such that targets cannot be seen much closer than about 2000 yards.

The transmitter is keyed and modulated by the units on the ground beside the trailer. These units were designed by Messrs. Vieweger, Moore, Noyes, and Marchetti. The keyer contains a 4100-cycle-per-second sinewave oscillator which establishes the basic pulse rate, and additional tubes which convert the sinewave into a series of sharp pulses. The transmitter tubes are operated at from 8000 to 15,000 volts. The ring circuit is coupled through a loop to the open-wire transmission line which conducts the pulses to the transmitting array, which consists of 16 half-wave radiators and 16 reflectors.

Each receiving array is connected to its respective receiver by two transmission lines taken from opposite ends of the array. Phasing stubs, in the center of each array, adjust the double-lobe pattern. The two terminations are fed to separate radio-frequency stages, which are switched on and off alternately by a rectangular wave of voltage applied to the cathode circuits at 1400 cycles per second. The plate circuits of the radio-frequency stages are connected together, so that in successive stages the double-lobe signals are amplified together in time sequence.

The radio-frequency signal is converted to intermediate frequency at 19.5 megacycles and amplified in four

intermediate-frequency stages which display a bandwidth of about 1 megacycle. The gain up to this point is about 20,000 times in voltage, or sufficient to reach the noise level of the input radio-frequency stages. Thereafter the signal is detected and amplified at video frequencies.

The receiver output is then conducted to its associated cathode-ray indicator, where, after further video amplification, the pulse signal is applied to the vertical deflection plates of the cathode-ray tube. The cathode-ray tube and its auxiliary circuits are similar to those of an ordinary test oscilloscope. The horizontal sweep is linear with time and occurs at a rate of 4100 sweeps per second, the rate being established by the sinewave oscillator in the keying unit. In addition, the horizontal sweeps are displaced slightly left and right in synchronism with the switching of the radio-frequency amplifiers in the receiver. Thereby two pulses are made to appear on the cathode-ray tube, one representing the signal from the left-hand lobe of the array, the other from the right-hand lobe. The resulting split image is equalized by the operator in orienting the array. The range indicator does not display a split image, since its function is to indicate simply the time difference between transmission and reception. The sweep circuit in this case is delayed by passing the sinewave from the keyer oscillator through a phase shifter. By adjusting the phase shift, the pulse can be moved across the screen until it falls under the reference hairline.

It may also be mentioned that the SCR-268 included a converter for the purpose of changing slant-range and elevation indications to altitude for use by the gun director, designed for us by Frankford Arsenal.

In all, the SCR-268 employs 110 vacuum tubes.

I would like to call your attention to the soldier operators of our developmental models. These soldiers became expert operators and their commander, First Lieutenant Cassevant, C.A.C., became an expert radar engineer. To them we owe much in military design.

Radars for Long-Range Warning

In 1938, work began on another radar, the SCR-270, to fulfill the requirement for long-distance warning against aircraft. By that time, the basic research at the Signal Corps Laboratories had revealed the means of accomplishing this objective. To obtain long range, the highest possible transmitter power and a large antenna, having high power gain, are required. In the receiver system, high gain in the antenna and the highest possible sensitivity are necessary. As further aids to long-range operation, the pulse energy (pulse amplitude times pulse length) should be high. This implies long pulse. Finally, in order to maintain high energy per unit of time on the cathode-ray tube screen, the spacing between pulses should be no longer than necessary for them to have time to travel to and from distant targets.

These requirements led, after many changes, to the following specifications: the transmitter power is between 30 and 100 kilowatts at a carrier frequency of 110 megacycles according to plate voltages used. The pulses are 15 to 40 microseconds long and transmitted at a rate of 625 per second. At 625 pulses per second, the interval between pulses permits detection out to 150 miles.

A single antenna array is used for both transmission and reception. This "duplex" operation permits the use of a single indicator and one operator. The array consists of 32 half-wave dipoles arranged 8 high and 4 wide, and a reflecting screen or alternately one that is 4 high by 8 wide, mounted on a metal tower.

The array is rotated in azimuth at a rate of about 5 revolutions per minute. The transmitted beam is 28 degrees wide and 11 degrees high between half-power points (or 11 degrees wide by 28 degrees high for alternate antenna). The beam rotates through its own width (28 degrees) in about a second, during which time some 625 radio pulses are sent out. Thus every point in space, surrounding the radar from the horizon to 11 degrees above the horizon, is continually "sprayed" with pulses. Aircraft in this region except for wave-interference spaces, and out to a distance of 100 miles or more, reflect visible echoes. By noting the direction and distance of particular echoes on successive turns of the array, the paths of the aircraft can be followed by plotting, at 12-second intervals, the point representing their position. The angular precision is, of course, poor compared to that of the SCR-268 since no lobe-switching is employed. But since the function of the radar was to warn, rather than to direct gun fire, this lack of precision can be tolerated.

Protection of the receiver is accomplished by the insertion of a spark gap in the receiver transmission line. This gap breaks down during the transmission of each pulse and thereby throws a short circuit across the receiver input. Between pulses, the gap is inactive and the received signal is passed to the receiver. In a later design, three such gaps are used to secure a more perfect short circuit during the transmitted signal.

In the design of the SCR-270, many improvements of an engineering nature were introduced. The transmitter consists of but two tubes (these tubes were developed for us by Westinghouse) operated at between 8000 and 15,000 volts plate potential. These are of the water-cooled variety and possess sufficient emission to reach a 350-kilowatt level from a pair of tubes when series keyed. In the production models, grid modulation was employed, and a 450TH tube, driven by a similar tube, served as the modulator. Under these conditions, 30 to 100 kilowatts is obtained from the transmitter. The receiver has a four-stage intermediate-frequency-amplifier preceded by an orbital-beam tube radio-frequency amplifier especially developed by RCA. The final video amplifier, feeding the cathode-ray indicator, is a beam tetrode. The indicator and range units are very similar to those of the SCR-268.

The reliable range of the SCR-270 is 120 miles on bomber aircraft targets, and about 75 miles on fighters.

The distance to the aircraft is indicated accurately to 2 miles and its direction to about 4 degrees.

A large number of modifications of the basic design were produced for special needs. The SCR-270 is trailer mounted and, with its associated trucks, can be moved over roads and set up quickly. Another series (the SCR-271), was produced for fixed location in permanent or semipermanent buildings. In all 788 radar sets SCR-270/271 were delivered by the prime contractor, the Westinghouse Electric and Manufacturing Company. After five years of use, the SCR-270 is still standard equipment, no radar set yet developed being able to replace it completely.

Fig. 28—Early models of both the SCR-270 and SCR-271 installed at Twin Lights, New Jersey.

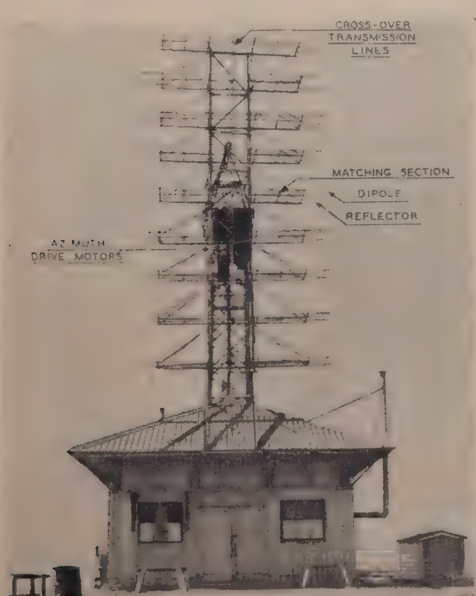


Fig. 29—An early SCR-271 installed at Twin Lights, New Jersey.

A view of a fixed and a mobile version of the SCR-270 is given in Fig. 28, and Fig. 29 is a close-up view of the fixed-station version of the SCR-270, known as the SCR-271. This mount was designed by the Blaw-Knox Company and the electrical fittings by Terpenning. Fig. 30 shows one of the late production models of the SCR-270. This antenna was designed by the Radio Corporation of America and is remarkable for bandwidth and absence of secondary lobes. The mount shown in the foreground was designed by Couse Laboratories and used in all production models. Fig. 31 shows the high-tower version of the SCR-271. The tower and mount were designed by Blaw-Knox Company.

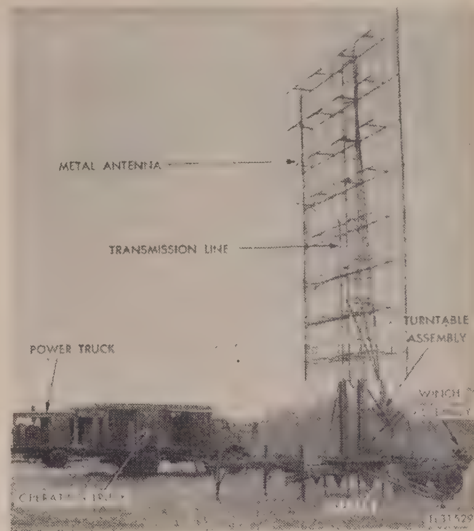


Fig. 30—Radio set SCR-270 set up for operation.

Fig. 31—Radio set SCR-271-D at the Evans Signal Laboratory, showing shelter for components and operating personnel, 100-foot tower, and antenna modified for plan-position-indicator display. Identification, friend, foe (IFF) antenna is shown at top center on radar antenna.

Components of the 270/271 are indicated in Fig. 32. The receiver and oscilloscope were designed by Mr. Moore; the production engineering and manufacturing were done by Radio Corporation of America. The transmitting tubes, also shown, are two spare transmitting tubes (WL-530), developed by Westinghouse. Figs. 33 and 34 are different views of the SCR-270/271

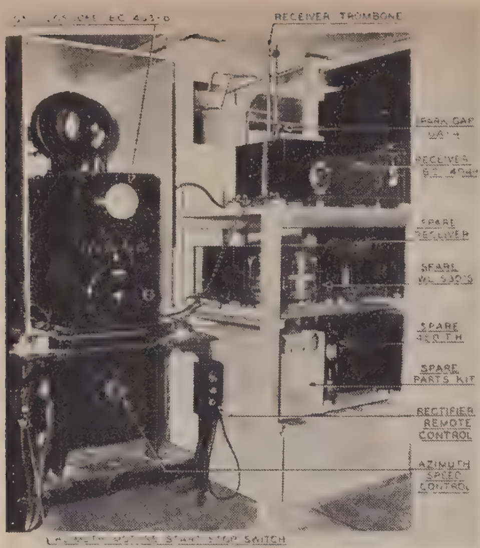


Fig. 32—Parts of an early SCR-270 installed in a K-30 truck.

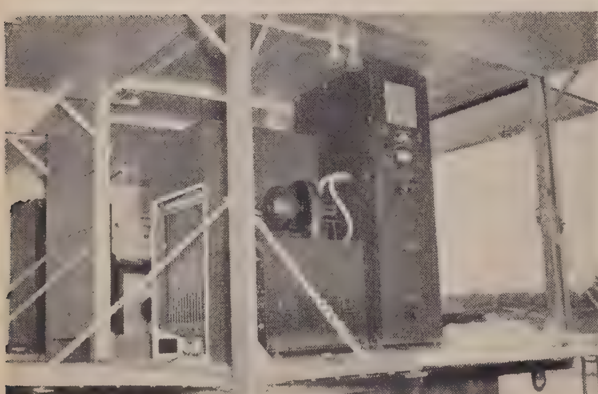


Fig. 33—Rear view of operating van of radio set SCR-270 showing transmitter, water cooler, and spare oscilloscope.

Fig. 34—Transmitter for radio set SCR-271-D.

transmitter. This transmitter was developed by Mr. Watson. The production engineering and manufacturing were done by Westinghouse.

Fig. 35 is a view of a complete SCR-270 as assembled for road transportation. Westinghouse was the prime contractor and had delivered 112 by Pearl Harbor day. From laboratory model to first delivery took only six months.

A most important modification is shown in Fig. 36. This was introduced after our entry in the war, and is a type of indicator known as the plan-position indicator

Fig. 35—Radio set SCR-270 packed, ready for transport.

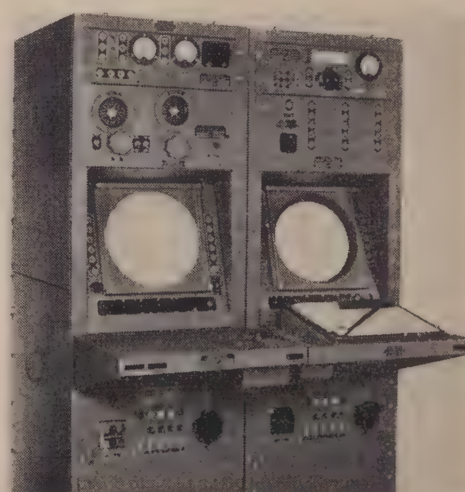


Fig. 36—Receiver indicator used with radio set SCR-270/271 modified for plan-position-indicator display.

(PPI). In this indicator, the cathode-ray beam is deflected radially outward from the center of the cathode-ray tube, with the transmission of the pulses. The cathode-ray beam starts at the center of the tube at the instant the pulse is transmitted and proceeds outward at constant speed, its position corresponding to the position of the pulse in space. When the pulse encounters a target and is reflected, the cathode-ray beam is brightened by intensity modulation and a spot of light appears on the screen, representing the target. Since the direction of the radial motion of the spot is controlled by the position of the array both the distance and the direction of the target are thus indicated. In effect, the screen represents a plan view of the area surrounding the radar, hence the name "plan-position indicator." The advantage of the

PPI is that it can display a large number of echoes simultaneously, whereas the simple indicator previously described (known as type A) is limited to one target at a time. This type of equipment is not unique to the SCR-270/271 but is rather standard to all modern radars. The chief difficulty in its development was obtaining a coating of proper persistence for the face of the oscilloscope. This particular equipment was designed by the General Electric Company.

Conclusion

This story would not be complete without giving due credit to Messrs. Trees, Rauh, Smith, Burtt, and Lewis, and their shop groups who corrected the errors of our engineers and who were responsible for much of the design.

This concludes the story of the 268 and 270/271, jointly developed by the Signal Corps Laboratories, Western Electric Company, Westinghouse Electric Corporation, Bell Laboratories, Radio Corporation of America, Blaw-Knox, Breeze, Couse Laboratories, Eitel-McCullough, Frankford Arsenal, General Electric Company, Le Roy, Fred M. Link, Rangertone, and Terpenning.

Frequency-Modulated Magnetic-Tape Transient Recorder*

HARRY B. SHAPER†

Summary—**A transient recorder having a frequency range of 0.02 to 1000 cycles per second with useful response up to 2000 cycles per second is discussed. The transient is recorded on a loop of magnetic tape and played back synchronously every 0.1 second on an oscilloscope screen. Thus, a steady image of the transient is obtained. Excellent signal-to-noise ratio (40 decibels) is obtained by the use of a 10-kilocycle carrier, which is frequency modulated. Each recording can be obliterated by simply pressing a button, with no material being consumed.**

I. Introduction

THE TRANSIENT behavior of phenomena has always eluded both direct measurement and mathematical analysis. Only in the simpler cases, where all the boundary conditions are known, can the mathematical methods be applied. If they can be applied, these methods are very powerful and illuminating. The great majority of transient phenomena are, however, not clearly defined and must be measured directly.

The instrument to be described is designed to facilitate the observation of transients. Amplifiers and oscilloscopes are now sufficiently well designed to reproduce faithfully the common transients which occur in shock vibrations, lighting, welding, switching, relays, etc. The difficulty is that the phenomenon occurs as a single flash on an oscilloscope, and must be photographed to be observed. Direct-inking pens are available, but these are limited in frequency range, and if high-speed transients are to be observed, then large quantities of paper are consumed. The photographic method is satisfactory but is extremely cumbersome and slow. This is especially true in laboratory investigations where a great number of trials are made and where adjustment of the apparatus is to be made.

A neat way to solve the problem is to record the transient and play it back so that the transient is repeated synchronously. Thus the transient appears as a steady-state signal on the oscilloscope. In this arrangement the signal may be observed conveniently. Such a medium is ideally available in a loop of magnetic tape or wire. No material is used up, and erasing the signal is easy and simple. A new signal can immediately be recorded and observed.

Fig. 1—The frequency-modulated magnetic-tape transient recorder.

This is the basis for the design of the instrument shown in Fig. 1. The instrument stands 23 inches high,

* Decimal classification: R365.3. Original manuscript received by the Institute, December 14, 1944; revised manuscript received, April 25, 1945. Presented, Cleveland Section, Cleveland, Ohio, November 30, 1944.

† The Brush Development Company, 3311 Perkins Avenue, Cleveland 14, Ohio.

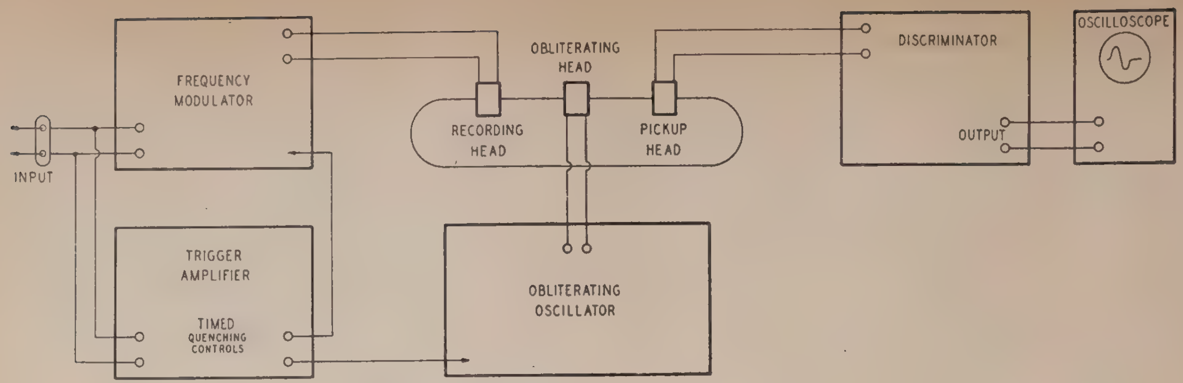


Fig. 2—Block diagram of transient recorder.

and is conveniently portable for laboratory use. It will play back transients of 0.7-volt peak signal with a 45-decibel signal-to-noise ratio. An attenuator in the input circuit permits the recording of 200-volt signals. The simple operation of the pressing of a button resets the instrument, and it is instantly ready for the next transient.

The method of operation is shown in Fig. 2. Before a signal is impressed on the unit, the carrier-frequency oscillator continuously records 10 kilocycles on the tape through the recorder head. The tape travels through the head at 25 feet per second and is 2.5 feet long; hence the tape makes one revolution every 0.1 second. Since the tape is driven by a synchronous motor, these revolutions occur synchronously. At the same time, the obliterating oscillator continuously obliterates the 10-kilocycle carrier. The obliterating head is driven at 30 kilocycles and obliterates by virtue of the fact that the tape is exposed to a number of wavelengths of decreasing strength. Thus, the tape leaves the obliterating head in a neutral state and no signal is picked up in the pickup head.

When a signal is impressed on the input terminals, it is divided into two channels, as shown. One modulates the carrier. While the frequency-modulated carrier is being recorded on the tape, the other channel drives a high-gain amplifier which sets off a pair of thyratron triggers. The first thyratron fires as soon as the transient signal reaches 20 millivolts. When the first thyratron fires, it quenches the obliterating oscillator before the modulated carrier reaches the obliterating head. The spatial relationship of the recording head and the obliterating head insures that all of the input transient is recorded. In fact, the base line (for zero signal) is 0.01 second long.

The first thyratron, by means of a resistance-capacitance delay, fires the second thyratron 0.1 second later. When this thyratron fires, it quenches the recording carrier. Thus, just as the part of the loop which retains the first part of the signal comes around to the recording head, the recording head is quenched and the overriding of the tail end by the initial part of the signal is prevented.

The frequency-modulated carrier is now picked up by the pickup head. This signal is repeated synchronously every 0.1 second. The signal is amplified, demodulated by a special discriminator circuit, and then applied to the vertical amplifier of a scope with a 10-cycle sweep and locked in. The original transient may now conveniently be viewed on the screen as a steady-state signal.

If it is desired to observe a second transient, the pressing of the reset button shown in Fig. 1 resets the thyratrons. The recording on the tape is obliterated and the carrier oscillator resumes the recording of the carrier. The unit is then ready for a new transient. If a record of the transient is desired it can be photographed, or the loop of tape may be removed and stored. The recorded signal on the steel tape may be considered an ideal and permanent record.

As is well known, every transient pulse may be resolved into a set of frequencies having a specific magnitude and phase relationship. In order to have a distortionless recording of these transients, both the magnitudes and the phase relationships of these components of frequency must be maintained over an infinite bandwidth.

In recording any particular transient, it is practicable to achieve this ideal for only those frequencies which carry most of the energy. For extremely sharp pulses, errors will be made in the steepness of the wave front. The reproduction will be less steep than the actual wave.

Thus, for a square pulse[1] of 0.1 second duration, a system must be capable of rising to full value in at least 0.01 second, or faster. In addition to rising rapidly to this value, the system must rise to full value and stay there. This means that the system must be "critically damped." In practice this is achieved by making certain that the system does not cut off sharply but is rounded at the cutoff frequency.

Furthermore, the linear phase-shift requirement provides that the low frequencies reach maximum amplitude in synchronism with the high frequencies which attain this maximum amplitude in the required 0.01 second. Thus, the top of the square wave is maintained flat by the low frequencies, once the high-frequency components have achieved this amplitude. If it is desired that the top of the square wave be maintained flat during the

[1] E. A. Guillemin, "*Communication Networks*," John Wiley and Son, New York, N. Y., 1935, vol. 2, p. 486.

pulse, then the lowest frequency required is roughly 10 times the 0.1 second, or 1 second. In addition, the low frequencies must not be time delayed relative to the high frequencies. If relative time shift takes place, then, even if the system reproduces these lows at correct amplitude, instead of the crest of the low frequencies appearing at the start of the transient, the sloping part of the cycles will appear, and distortion occurs.

In practice, then, the actual time delay of the high frequencies versus the lows, (due to the phase shift being nonlinear) must not exceed 1/20 of the time of the duration of the pulse.

The above conditions are the requirements for even crude reproduction of this particular transient of a square-wave pulse. They serve to show the need for great care and precision in the reproduction of transients.

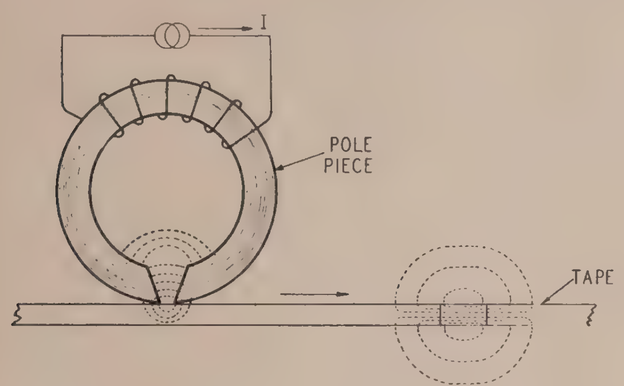


Fig 3—A method of recording signals on a magnetic tape.

In order to understand the reason for the use of frequency modulation rather than amplitude modulation in this recorder, it is first necessary to examine the method of recording and reproducing signals on a magnetic-tape loop in an elementary way.

If a current I flows through the coil of Fig. 3, the induced magnetomotive force

$$MMF = 0.4\pi nI \text{ gilberts} \tag{1}$$

where n is the number of turns in the coil. The resultant flux which flows is

$$Q = MMF[(R_I + R_a)/R_IR_a] = MMF/R \tag{2}$$

where R, the reluctance of the system, is given by the combined reluctance of the tape itself R_I and the reluctance of the air leakage path R_a. Since the soft iron core has a permeability 1000 times that of the air, practically the entire generated magnetomotive force of $0.4\pi nI$ gilberts will be applied across the exposed piece of tape. The flux through the tape then will be governed by the properties of the tape. The characteristic hysteresis loop of the tape material is similar to that of most magnetic materials, and is shown in Fig. 4. The co-ordinates represent the flux through the tape and the magnetomotive force across the ends of the tape element. Since the tape is fully demagnetized by the obliterating head, it arrives at the recording head in the neutral state shown at 0. Due to the magnetomotive force, the tape is magnetized along the characteristic curve and shifts from the origin to the position M, depending on the strength of the field. When it leaves the influence of the magnetomotive force of the pole pieces, a new set of flux conditions obtain as shown in Fig. 3. This leakage flux causes a magnetomotive force drop to take place in the magnetized section of the tape, and the magnetic condition of the tape varies along a minor hysteresis loop to the position P.

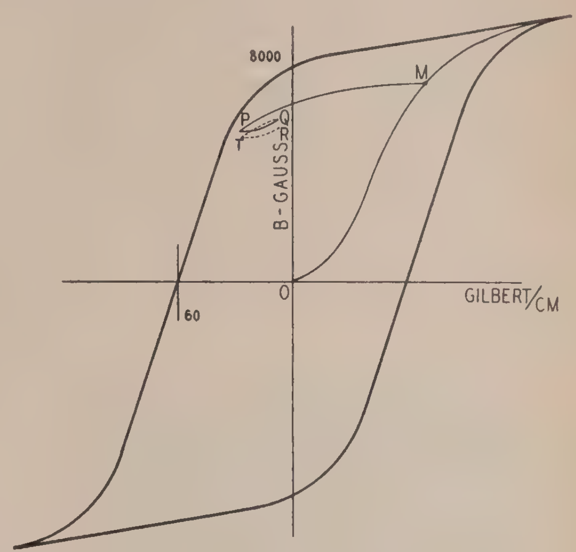


Fig. 4—Characteristic hysteresis loop, magnetic-tape material.

When the tape then runs under the playback-head pole pieces, the loop runs to the short-circuit magnetic position Q. The flux Q then induces a voltage

$$E \doteq NdQ/dt \times 10^{-8} \text{ volts.} \tag{3}$$

Where N is the number of turns in the pickup and (dQ/dt) is governed by the speed of the moving-tape element. On leaving the short-circuiting pole pieces of the pickup, the tape magnetic condition obtains at T, due to the fact that the flux in the air medium again assumes the flux-leakage paths shown in Fig. 3. After a number of

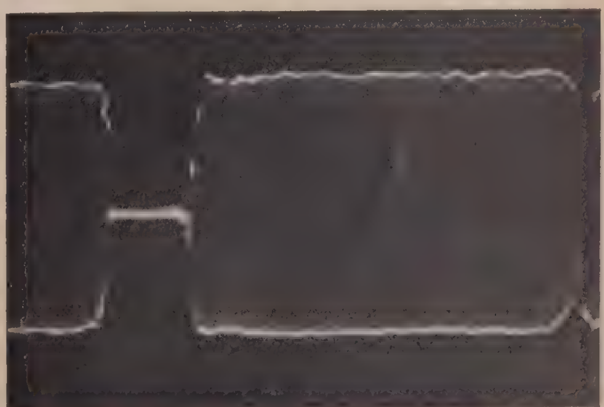


Fig. 5—A 10-kilocycle recorded signal.

playback short circuits, a minor hysteresis loop is asymptotically established at R and from then on the cycles repeat themselves.

Thus, we see that the linearity of the reproduced signal is governed by the linear relationship between the induced magnetomotive force at M and the resulting induced flux R in the playback head. While this linearity has been striven for by careful manufacture of the tape, it has been found that different sections of the tape have slightly different properties due to rolling, handling, and composition variations in the material.

If it is desired to reproduce direct-current signals, it is necessary to use some form of modulated carrier due to the differentiation which takes place in the pickup head, as shown by (3). For if direct-current were recorded by the recording head according to (2), the playback voltage would be zero, since the flux in the pickup head would be constant as shown by (2).

A typical 10-kilocycle recorded signal is shown in Fig. 5. Note the small changes in amplitude of the carrier signal along the tape. The figure shows a larger

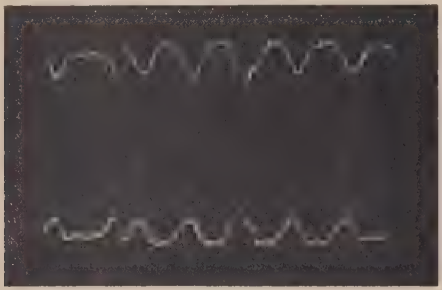


Fig. 6—Showing amplitude modulations due to cutoff of higher frequencies.

change in recorded signal at one point. This is due to the change in properties at the tape joint. This is present at the point where the tape is welded in order to make a continuous loop. The region of zero carrier is shown at the left where the obliterating head has removed the carrier and the tape is in a neutral state.

If this 10-kilocycle signal is amplitude modulated, these deviations in carrier amplitude will appear as background noise, and the ultimate signal-to-noise ratio will be governed by the inherent lack of uniformity in the tape.

However, with the use of frequency-modulation signals, the error in the amplitude of the signal is avoided, since the signal on the tape is retained only by the frequency and its rate of change.[2] Amplitude variations in the signal do not enter into the playback process, and, in fact, are carefully removed by severe limiting of the signal before the operation of the discriminator takes place. Thus, even the violent amplitude change at the tape joint is removed. The limitation in the signal-to-noise ratio for frequency modulation occurs with those factors which involve time errors. Thus, if the tape does not run at uniform speed or "flutters," then distortion occurs in the playback.

A second-order error occurs if the tape is very nonuniform. Then the recording flux jumps ahead or is retarded in the recording and playback process as it seeks the more permeable part of the tape.

II. Frequency Modulator

The selection of the circuits for the modulator is based on the requirements and limitations of the tape. We set ourselves the requirement that transients of 0.1-second duration were to be measured. A convenient size of loop of approximately 2.5 feet in length was selected. This meant that the loop had to travel at about 25 feet per second. The wavelength of the carrier on the tape then is $\lambda = (25 \times 12)/10{,}000 = 0.030$ inch.

The pole-piece gap on the pickup head could then be 0.007 inch which would be less than $\frac{1}{4}$ wavelength. If the pole-piece gap becomes the order of 1 wavelength, then cutoff takes place due to the neutralizing action on the head from the out-of-phase components of the cycle which are recorded on the tape. The highest carrier frequency is selected, which, when frequency modulated, would not have the higher frequencies cut off. Actually the heads do cut the higher frequencies with resultant amplitude modulations (see Fig. 6). The amplitude modulations are, however, eliminated in the limiter. Thus, operation is attained at the highest frequency before cutoff takes place.

The per cent modulation or modulation factor is made as wide as possible consistent with linear frequency modulation and demodulation. The wider the swing, the less the influence of "flutter."

Once the carrier frequency is determined, then the upper limit of modulation frequency which can be reproduced with reasonable accuracy is set. The limitations are set by the selectivity of the filters and the time constants of the circuits employed.

There are several methods which are practical for use as frequency modulators. Each of them has its own particular advantages and disadvantages. In the laboratory model two of the types were tried and neither of them was found to be ideal. Consequently, a new direct-modulator scheme has been developed.

The advantage of using a beat-oscillator scheme for a frequency modulator lies in the fact that, by raising the frequencies of the beat oscillators to approximately 50 times that of the carrier frequency, excellent linearity can be obtained. However, the effective drift in the carrier frequency is accordingly excessive. This can be controlled by regulating the power supply and using a set of push-pull modulator tubes, one of the tubes acting as a variable inductance and the other acting as a variable capacitance.[3] The above system is satisfactory but requires an excessive number of

[2] August Hund, "Frequency Modulation," McGraw-Hill Book Company, New York 18, N. Y., 1943, chapter 1.

[3] See p. 175 of footnote reference 2.

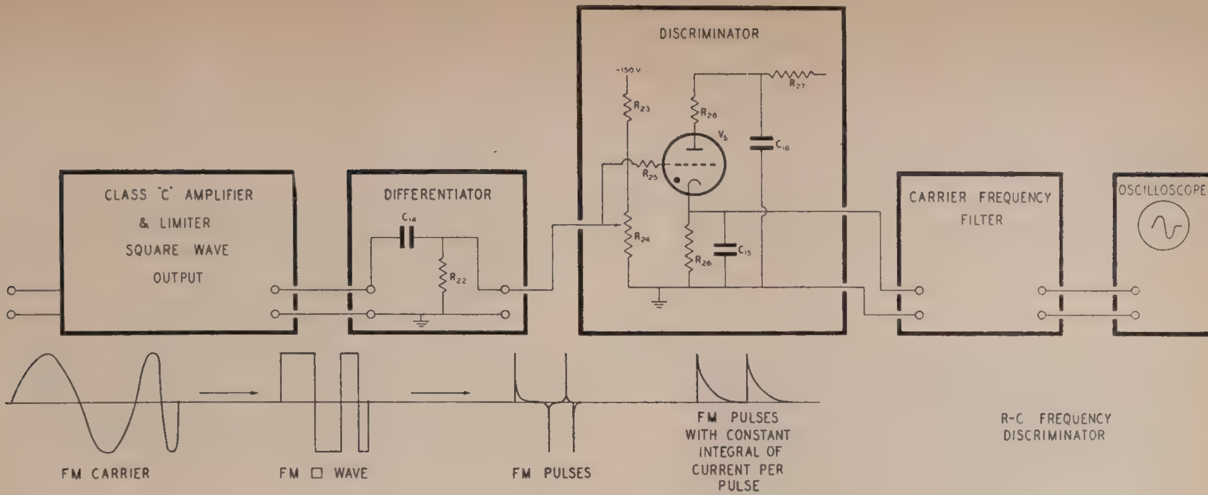


Fig. 7—Block diagram of playback mechanism.

R_{22} = 25,000 ohms	R_{25} = 25,000 ohms	C_{14} = 10 micromicrofarads
R_{23} = 200,000 ohms	R_{26} = 15,000 ohms	C_{15} = 0.003 microfarad
R_{24} = 25,000 ohms	R_{27} = 250,000 ohms	C_{16} = 50 micromicrofarads

tubes and circuit components, which rules it out for use in a portable unit.

A method of modulating a resistance-capacitance oscillator directly has been reported.[4] Here, a variable-impedance tube places a variable capacitance or resistance across the resistance-capacitance feedback path and causes frequency modulation in accordance with changes in the grid voltage. When large frequency swings are obtained by this method, they are accompanied by large amplitude modulations. When the amplitude modulation is corrected by the use of a limiter, an excessive number of components is used.

This amplitude modulation is avoided in an alternative resistance-capacitance frequency-modulation system.[5] Here a ladder-type resistance-capacitance network is used with triode plates acting as variable resistors in accordance with the grid bias. When wide swings are obtained they are accompanied by relatively small amplitude modulations. The disadvantage in using this method is the difficulty in obtaining linearity over wide swings. Linearity is achieved by employing the nonlinear plate resistance of the triode (with grid voltage) to compensate for the nonlinear, but opposite, swing in oscillator frequency with change in the two resistances to ground.

III. Frequency Discriminator

The method of playback to the oscilloscope is shown in Fig. 7. The frequency-modulated carrier is picked up by the head and applied at the terminals of a high-gain class C amplifier. Use is made of class C so that the obliterated portion of the tape will have its noise-level signal suppressed and no signal will be fed to the discriminator. Thus, a clean indication of the start and stop of the transient is obtained.

The amplifier overloads, and the signal picked up is strongly limited. Thus, the signal coming out of the amplifier is a frequency-modulated carrier wave with square tops. This signal is then differentiated by C_{14} and R_{22}, since the impedance of the capacitor at 10 kilocycles is more than an order of magnitude larger than the resistance. The square waves are thus converted into pulses.

The "discriminator" used is a novel and extremely simple one. It converts each pulse of the carrier frequency into a current pulse whose integral is a constant. Thus, the more cycles which occur in any one second, the higher will be the integrated total current over the interval; in others words, the higher the frequency, the higher the output current. This relationship is achieved

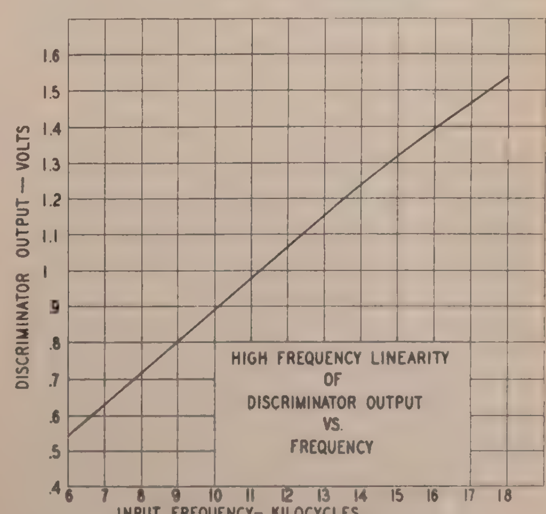


Fig. 8—Showing linearity of discriminator characteristic.

[4] C.-K. Chang, "A frequency-modulated resistance-capacitance oscillator," Proc. IRE, vol. 31, pp. 22–25; January, 1943.
[5] Maurice Artzt, "Frequency modulation of resistance-capacitance oscillators," Proc. IRE, vol. 32, pp. 409–414; July, 1944.

in a linear manner by the use of the thyratron V_5 in Fig. 7. R_{27} and C_{16} coupled to the thyratron, act in the usual relaxation manner. However, by means of the potentiometer R_{24}, the bias on the thyratron is set so that relaxation oscillation is prevented. When one of the carrier-frequency pulses is applied to the grid, the capacitor C_{16} fires through the tube and discharges itself completely into C_{15}, since C_{15} is much larger than C_{16}. The sharp triggering pulse on the grid is now over, and the capacitor C_{16} charges up to the B-supply voltage.

Fig. 9—Oscillogram showing shape of the pulses delivered at the cathode of the thyratron for a 7-kilocycle frequency.

The grid regains control, since it is biased to cutoff, and the tube is ready for the next triggering pulse. The coulombs delivered to C_{15} are controlled by the B-supply voltage and the capacitance of C_{16}. The B-supply is regulated and C_{16} is an air capacitor, hence the constancy of the coulombs delivered per pulse is assured.

Fig. 8 shows the linear relationship between input frequency to the discriminator and the output volts as measured at the cathode of the thyratron. It is interesting to note that this type of discriminator is untuned and is linear down to direct current. For direct current there is zero signal since no pulses arrive at the thyratron. For one cycle per second, one pulse per second is delivered by the thyratron; for two cycles per second, two pulses per second are delivered, etc.

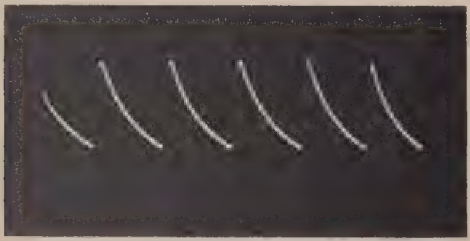


Fig. 10—Oscillogram of 14-kilocycle carrier. Note that the tails are not fully decayed before interruption.

Fig. 9 shows the shape of the pulses delivered at the cathode of the thyratron. The oscillogram is taken for a 7-kilocycle frequency. The sharp rise is the discharge of the thyratron and the decay is the usual one for a resistance-capacitance circuit. The tail of the decay curve limits the highest frequency which can be linearly discriminated. If the tube fired in the middle of the curve, the capacitor C_{16} would not be fully charged and therefore would not deliver the full number of coulombs. The curve in Fig. 8 starts to round off at about 14 kilocycles because of this. Thus, the two resistance-capacitance circuits used in conjunction with the thyratron determine the upper limit of linear discrimination. Fig. 10 shows the oscillogram of a 14-kilocycle carrier where the tails have not fully decayed before interruption.

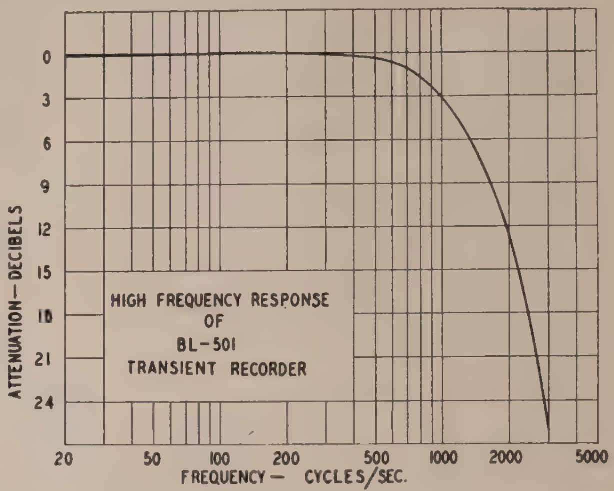


Fig. 11—High-frequency response characteristic of transient recorder.

Several circuits have been devised which eliminate the characteristic long tails. These were not used since higher frequencies are not required in this system, and also because these circuits called for the use of some additional components. In the event that wider frequency-range transients were to be recorded, there would have been no choice and the faster circuits would have had to be employed.

The output of the discriminator thus contains a set of constant integral pulses which appear at the frequency-modulated rate. A carefully designed low-pass filter eliminates the carrier, leaving the original transient. The filter must have linear phase shift and must be rounded at the cutoff to prevent overshoot. The frequency-response characteristic of this filter essentially governs the frequency-response characteristic of the entire recorder and is shown in Fig. 11. At 1000 cycles the filter rounds off smoothly. The filter as designed had an m-derived section with infinite cutoff at 6 kilocycles and

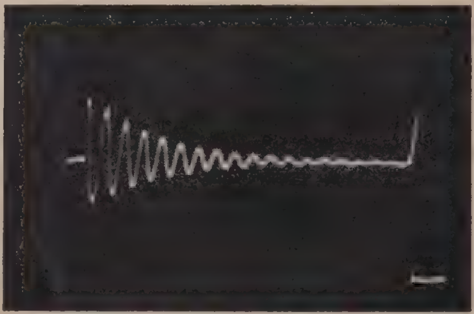


Fig. 12—Showing the decay of a typical oscillation.

a constant K section. Each section of the filter is separately damped to insure the smooth cutoff characteristic with no overshoot. Only the high end of the frequency-response characteristic is shown. The system would go to direct current except for an input capacitor-resistor circuit which starts to cut the low frequencies at 0.01 cycle per second.

Fig. 13—Transient caused by discharging a capacitor into a resistor.

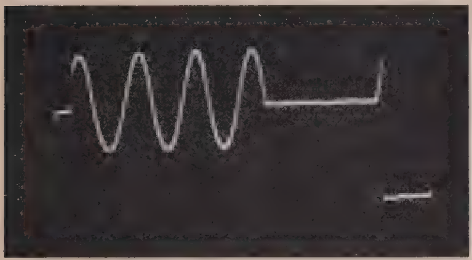


Fig. 14—Transient caused by switching.

Fig. 15—Transient caused by spoken syllable "aw."

In this manner the synchronously repeated transient is applied to the terminals of the oscilloscope, and a steady-state view of the transient is obtained.

IV. Conclusion

A set of typical transients is shown, to which the instrument is applicable. The graphs are composed of three parts: a 0.01-second zero-signal base line; the transient proper; and an obliterated carrier-signal section. The time base is always 0.1 second. Using this as a base, the frequency components of the transients may be calculated.

Fig. 12 shows a typical oscillation decay. A battery was simply switched into a series inductance-capacitance circuit and the voltage across the coil was applied at the input terminals of the recorder.

Fig. 13 shows the decay of a charged capacitor into a resistance. The voltage across the resistance was applied across the input terminals of the recorder. Note the sharp rise from the base line and the lack of overshoot at the peak. This is actually one of the more difficult transients to record, in spite of the simplicity of the circuit which yields this transient.

Fig. 14 shows three and one-half cycles applied to the input of the recorder and then cut out by a switch. This

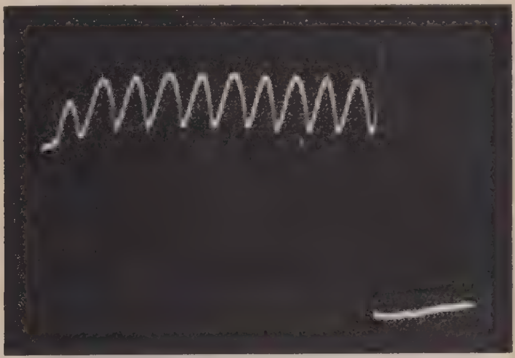


Fig. 16—Transient recording of the start of a single-element fluorescent lamp.

Fig. 17—Transient recording of the start of two fluorescent lamps.

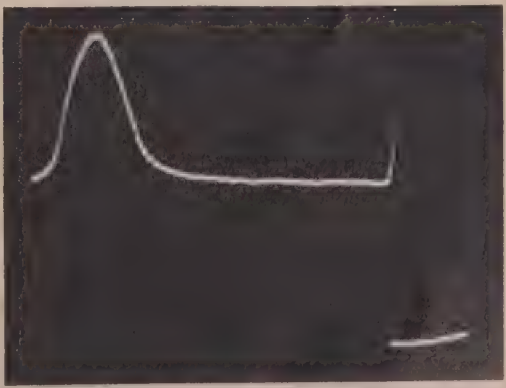


Fig. 18—Showing the timing of a photo-flash lamp.

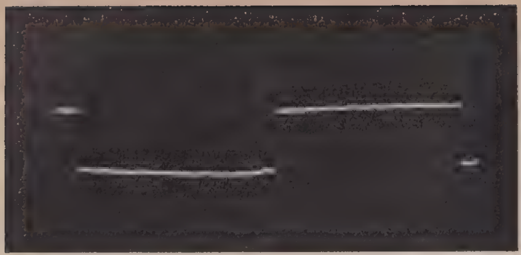


Fig. 19—Reproduction of 10-cycle-per-second square wave.

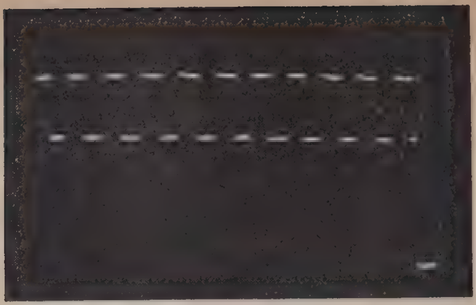


Fig. 20—Showing the response of transient recorder to a 100-cycle-per-second square wave.

is the type of transient which obtains in an ideal resistance weld.

Fig. 15 shows the build-up time of the transient caused by the spoken syllable "-aw-" as in the word "bawl." A microphone and amplifier were connected to the recorder, and the transient recorded includes the start of speech to the steady-state sound. With the same hookup, the low-frequency transients of loudspeakers could be observed.

Fig. 16 shows the transient recording of the start of a single-element fluorescent lamp. A photocell was placed directly across the input of the recorder and the lamp was simply turned on. Note the lower efficiency of the initial cycles. Note also that from the base line the flicker is about 75 per cent.

Fig. 17 shows a repeat of Fig. 16 except that two lamps are used. Since the lamps do not start simultaneously, only the steady-state light is shown. Note that the flicker as shown relative to the base line is now 30 per cent. Included in the recording is the step impulse of the direct-current light.

Fig. 18 shows the timing of a General Electric No. 5 photo-flash lamp with the same photocell connections as above.

Fig. 19 shows the reproduction of a 10-cycle-per-second square wave. Actually, the response of the system is shown to a step impulse since the 10-cycle frequency is slow enough to cover the tape loop in one cycle.

Fig. 20 shows the response of the system to a 100-cycle-per-second square wave.

Glossary of Disk-Recording Terms*

HOWARD A. CHINN†, SENIOR MEMBER, I.R.E.

Introduction

During 1941 the Engineering Committee of the National Association of Broadcasters formed a Recording and Reproducing Standards Committee for the purpose of formulating standards for electrical transcription and recordings made for radio broadcasting. Sixteen technical standards and good engineering practices were adopted[1] and are now followed by the industry.

A number of other items were under study by subcommittees and one additional item was completed before the war interrupted all activity in this field. The item in question is a glossary of disk-recording terms prepared by Subcommittee III and approved by the Executive Committee. This material has never been published, however, and in view of the potential postwar interest in recording, it seems opportune that the glossary be made available generally, particularly since it is not believed that material of this type exists elsewhere in the literature.

The members of the subcommittee which prepared the glossary consisted of H. A. Chinn, Columbia Broadcasting System, Inc., Chairman; E. J. Content, WOR; C. Lauda, World Broadcasting System; and G. E. Stewart, National Broadcasting Company.

Glossary of Terms Pertaining to Constant-Rotational-Speed Disk Recording

Abrasive: The grinding material sometimes incorporated in record stock for the purpose of shaping the needle point to fit the groove properly.

Acetate disks: Various acetate compounds used for solid and laminated (which see) disks. The term is often erroneously used to describe cellulose-nitrate discs (which see).

Advance ball: A rounded support (often sapphire) attached to the recording head which rides on the discs to maintain a uniform mean depth of cut by correcting for small variations in the plane of the disc surface.

Angle of groove: The angle from wall to wall of an unmodulated groove in a radial plane perpendicular to the surface of the disk.

* Decimal classification: R030×621.385.971. Original manuscript received by the Institute, February 23, 1945; revised manuscript received, May 18, 1945.

† Columbia Broadcasting System, Inc., New York, N. Y.

[1] L. C. Smeby, "Recording and reproducing standards," Proc. I.R.E., vol. 30, pp. 355–356; August, 1942.

Backed stampers: A thin, metal matrix (which see) which is attached to a backing material, generally a metal sheet 1/8 inch to 1/4 inch thick.

Binder: A resinous material which causes the various materials of a record compound to adhere to one another.

Biscuit: A small slab of the stock material, from which records are pressed, as it is prepared for use in the presses.

Blank groove: A groove upon which no modulation is inscribed.

Burnishing surface (of cutting stylus): The portion of the cutting stylus directly behind the cutting edge which smooths the groove.

Burnishing tool: The stylus sometimes used to smooth the groove of a recording.

Cake wax: A thick disk of wax (which see) upon which an original recording is inscribed.

Capacitor pickup: A phonograph pickup which depends for its operation upon the variation of its capacitance.

Carbon-contact pickup: A phonograph pickup which depends for its operation upon the variation in the resistance of carbon contacts.

Cellulose-nitrate disks: See lacquer disks.

Center hole: The hole in the center of the record, which fits the center pin of the turntable.

Center pin: The shaft protruding from the center of the turntable used for centering the record.

Chip: The material removed from the disk by the recording stylus in cutting the groove.

Christmas-tree pattern: A term sometimes used in referring to the optical pattern (which see).

Condenser pickup: See capacitor pickup.

Constant amplitude: Recordings wherein all frequencies of the same intensity are inscribed at the same amplitude.

Constant velocity: Recordings wherein frequencies of a given intensity are inscribed with the same maximum velocity of the cutting stylus.

Core: The central layer or basic support of certain types of laminated disks. (See laminated and lacquer disks.)

Crossover frequency: See transition frequency.

Crossover spiral: Same as spread groove (which see).

Crystal pickup: A pickup which depends for its operation on the piezoelectric effect of certain (generally Rochelle salt) crystals.

Cutter: An electromechanical transducer which transforms electric energy into mechanical motion which is inscribed into the original record by the cutting stylus. Also known as recording head.

Cut double (triple, etc.): Make two (three, etc.) original recordings simultaneously.

Cutting stylus: The cutting tool which cuts the groove into the original record.

Damping: See mechanical damping.

Drive pin: A pin similar to the center pin, but located to one side thereof, which is sometimes used to prevent the record from slipping on the turntable.

Drive-pin hole: A hole in the record which fits over the turntable drive pin.

Dubbing (in a cutting stylus): Same as burnishing surface (which see).

Dubbing (in recording): A recording made by re-recording from one or more records.

Dulling: Forming the burnishing surface of the cutting stylus.

Duping: To make duplicates by re-recording.

Dynamic pickup: A phonograph pickup in which the electrical output results from the motion of a conductor in a magnetic field.

Eccentric circle: A blank, locked groove (which see) whose center is other than that of the record (generally used in connection with mechanical control of phonographs).

Eccentricity: The eccentricity of the recording spiral with respect to the record center hole.

Fast spiral: A blank, spiral groove having a pitch that is much greater than that of the recorded grooves.

Feedback cutter: A cutter provided with a feedback circuit (separate from the driving circuit) in which a voltage, for inverse feedback to the driving amplifier, is induced by the movement of the cutting stylus.

Filler: The bulk material of a record compound as distinguished from the binder (which see).

Flowed-wax platter: Disk base (usually metal) upon which wax is flowed.

Flutter: Frequency modulation caused by spurious variations in groove velocity.

Frequency record: A record upon which have been recorded various frequencies throughout the desired frequency spectrum.

Groove: The track cut in the record by the stylus.

Groove contour: The shape of the groove in a radial plane perpendicular to the surface of the record.

Groove speed: See groove velocity.

Groove velocity: The linear velocity of the groove with respect to the stylus.

Grouping: Nonuniform spacing between grooves.

Guard circle: An inner concentric groove inscribed on a record to prevent reproducer from being damaged by being thrown to the center of the record.

Hill-and-dale recording: See vertical recording.

Hot plate: A heated table used for (a) softening the biscuits of record material prior to placing them in the press or (b) making flowed waxes.

Instantaneous recording: A recording which may be used without further processing.

Label: The identification markings on paper or similar material, at the center of the record.

Lacquer disks: Disks, usually of metal, glass, or paper, which are coated with a lacquer compound (often containing cellulose nitrate) and used either for "instantaneous" recordings or lacquer masters.

Lacquer master: A term improperly applied to a "lacquer original" (which see).

Lacquer original: An original recording on a lacquer disk which is intended to be used for the making of a metal master.

Laminated record: A disk composed of several layers of material. Normally used with one thin facer on each side of a core.

Land: The record surface between two grooves.

Lateral compliance: The ability of a reproducing stylus to move laterally with respect to the record groove while in the reproducing position in a record.

Lateral recording: A recording in which the groove modulation is in the plane of the record and along a radius.

Lead screw: The threaded rod which leads the cutter or reproducer across the surface of the disc.

Lead-in spiral: A blank, spiral groove at the beginning of a record, generally having a pitch that is much greater than that of the recorded grooves.

Locked groove: A concentric, blank groove at the end of modulated grooves whose function is to prevent further travel of the reproducer.

Magnetic pickup: A reproducer employing an armature placed in a magnetic field and coupled mechanically to the reproducing stylus. An electric potential is generated in a coil placed in this field when the stylus is actuated by the modulated groove of a record.

Master: The negative produced from the original recording (which see).

Master stamper: A master used as a stamper to make pressings.

Matrix: The negative from which duplicate records are molded. (See also stamper.)

Mechanical damping: The mechanical resistance which is generally associated with the moving parts of a cutter or a reproducer.

Metal master: The metal negative produced directly from the original recording.

Metal negative: Same as metal master (which see).

Metal mold: Same as matrix (which see).

Mother: A positive produced directly from the metal master or negative.

Needle (reproducing needle): A replaceable reproducing stylus (which see).

Needle drag: Same as stylus drag (which see).

Needle pressure: Same as stylus pressure (which see).

Optical pattern: The pattern which is observed when the surface of a record is illuminated by a beam of parallel light.

Orange peel: Mottled surface of a defective disc having an appearance similar to the skin of an orange.

Original recording: See lacquer original and wax original.

Overcutting: Excessive level in recording to an extent that one groove cuts through into an adjacent one.

Pickup: A mechanicoelectrical transducer which is actuated by the undulations of the record groove and transforms this mechanical energy into electrical energy.

Pinch effect: A pinching, or in some cases a lifting of the reproducing stylus, twice each cycle in the reproduction of lateral recordings, caused by the recording stylus cutting a narrower groove when moving across the record while swinging from a negative to a positive peak.

Playback: An expression used to denote the immediate reproduction of a recording.

Poid: The curve that the center of a sphere traces when the surface of the sphere is rolling along a sine wave.

Postemphasis: The complement in reproduction of pre-emphasis (which see).

Pre-emphasis: A method of recording whereby the relative recorded level of some frequencies is increased with respect to other frequencies.

Pressing: A record produced in a record-molding machine from a matrix or stamper.

Processing: Making the master, mother, and matrix (which see).

Recording head: Same as cutter (which see).

Re-recording: A recording made from the reproduction of a recording. (See also dubbing.)

Reference recording: Recording of a program or other material made for the purpose of checking same.

Reproducing stylus: The "needle" or jewel which follows the undulations in the record groove and transmits the mechanical motion thus derived to the pickup mechanism.

Rumble: Low-frequency vibration mechanically transmitted to the recording or reproducing turntable and superimposed on the reproduction.

Safety: A second recording, made simultaneously with the original, to be used for duplication should the original be damaged.

Shaving: Process of removing material from a wax disc of recording material to obtain a plane surface.

Shell or shell stamper: A thin metal matrix (generally 0.015 to 0.020 inch thick).

Spew: The excess record material which is ejected from the record press in the manufacture of pressed records.

Spread groove: A groove, with greater than normal pitch, cut between recordings of short-time duration, thus separating the recorded material into bands while still enabling the reproducing stylus to travel from one band to the next.

Sputtering: A process sometimes used in the production of the metal master, wherein the wax or lacquer original is coated with an electrical conducting layer by means of an electrical discharge in a vacuum. Sometimes called cathode sputtering.

Stamper: A negative (generally made of metal) produced from the mother (which see) and from which the finished pressings are molded. (See also matrix.)

Stylus drag: The expression used to denote the effect

of the friction between the record surface and the reproducing stylus.

Stylus force: Effective weight of reproducer or force in vertical direction on stylus when it is in operating position.

Stylus pressure: Term sometimes erroneously used to denote effective weight of reproducer or stylus force (which see).

Stylus weight: Actually stylus force (which see).

Surface noise: The noise reproduced in playing a record due to rough particles in the record material and/or irregularities in the walls of the groove left by the cutting stylus.

Throw-out spiral: A blank spiral groove at the end of a recording, generally at a pitch that is much greater than that of the recorded grooves.

Throw-out tail: End of throw-out spiral (which see).

Tracing distortion: A harmonic distortion introduced in the reproduction of records because of the fact that the curve traced by the center of the tip of the reproducing stylus is not an exact replica of the modulated groove. For example, in the case of a sine-wave modulation in vertical recording, the curve traced by the center of the tip of a stylus is a "poid" (which see).

Tracking error: The angle (in a lateral recording) between the vertical plane containing the vibration axis of the mechanical system of the reproducer and a vertical plane containing the tangent to the record groove.

Transition frequency: The frequency at which the change-over from constant-amplitude recording to constant-velocity recording takes place.

Translation loss: The loss in high-frequency reproduction which occurs as the groove velocity decreases.

Turnover frequency: Same as transition frequency (which see).

Vertical compliance: The ability of a reproducing stylus to move in a vertical direction while in the reproducing position on a record.

Vertical recording (hill-and-dale recording): A recording wherein the groove modulation is in a plane tangent to the groove and normal to the surface of the record.

Vertical stylus force: See stylus force.

Wax: A blend of waxes with metallic soaps (also see cake wax).

Wax master: A term improperly applied to a "wax original" (which see).

Wax original: An original recording on a wax surface for the purpose cf making a metal master.

William (or willy): A negative produced from a mother to produce still another mother.

Wow: A low-frequency flutter (which see).

The Servo Problem as a Transmission Problem*

ENOCH B. FERRELL†, SENIOR MEMBER, I.R.E.

Summary—**The purpose of a servo is to reproduce a signal at a place or power level or form different from the original signal, but under its control. It is therefore a signal-transmitting system. It uses negative feedback to minimize noise and distortion, which the servo designer usually calls error. It uses mechanical and thermal circuit elements as well as electrical circuit elements, but the problems of circuit design are the same.**

The methods of Nyquist and Bode, which have proved so useful in the design of electrical feedback amplifiers, are equally useful in the design of servo systems. They encourage the determination of the significant constants of the system by experimental means involving steady-state amplitude measurements.

Introduction

THE PURPOSE of this paper is twofold: first, to point out that many mechanical problems, servo problems in particular, can be handled by the circuit-analysis and circuit-design techniques of electrical engineering which the communication and radio engineers have developed to such a high degree; second, to encourage those faced with servo problems to study either the original papers or Nyquist and Bode on this subject or some of the more recent literature in textbooks or handbooks.

* Decimal classification: 621. Original manuscript received by the Institute, March 26, 1945.
† Bell Telephone Laboratories, Inc., New York, N. Y.

The purpose of a servo system is to reproduce a signal at a place, or at a power level, or in a form that is different from the original signal, but is under its control. It is, therefore, a signal-transmitting system. It uses negative feedback to minimize noise and distortion. All this can be said just as well about a telephone repeater, a public-address amplifier or an intermediate-frequency amplifier. The servo signal is usually mechanical. The circuit elements are motors, gears, or thermostats. The noise and distortion are called error. But the basic problems of stability, bandwidth, and linearity are just the same. And the simple, straightforward design technique that is based on circuit theory, and that has been used so successfully with negative-feedback amplifiers, is just as simple and straightforward when used with servo systems.

Let us talk in terms of an example. Suppose we have an input shaft. It may be a steering wheel, or a tuning knob, or the shank of a thermostat. It takes on various positions and motions. We want to reproduce those positions and motions at an output shaft, which may be a rudder, a tuning capacitor, or a fuel valve.

This varying motion is a signal in just the same sense that a varying electrical voltage or current is a signal. It is a function of time. We may study it by the method of

transients; or we may analyze it by Fourier's method and talk about its spectrum.

We find immediately one important difference between this signal and those we are accustomed to handle in speech and television. It is a low-frequency signal. It nearly always includes direct current. (Why not borrow the phrase, direct current, and apply it to an unvarying mechanical velocity!) Its spectrum seldom extends above a very few cycles per second. Often the highest frequency of interest is even less than a tenth of a cycle per second.

In our example, let us connect the input shaft to the brush of a potentiometer. If the potentiometer is excited by a battery, our output will be an electromotive force that is proportional to the displacement of the input shaft. Let us devise a name for the constant of this proportionality. We commonly call the ratio of a mechanical force to a mechanical displacement a stiffness. We commonly call the ratio of an electrical force to an electrical displacement an elastance, the reciprocal of capacitance. The usual symbol for elastance is S, which recognizes its similarity to stiffness. Let us combine these concepts of stiffness and elastance, and apply them to the ratio of an electrical force to a mechanical displacement. We will say the potentiometer has a stiffness. Since the force and displacement are measured on opposite ends of this circuit element, we call it a mutual stiffness S_m.

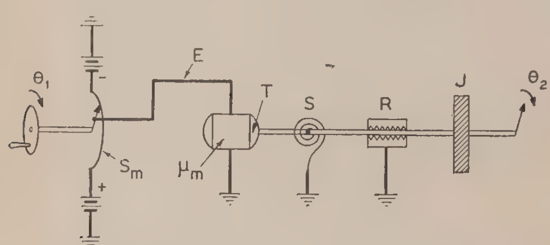


Fig. 1—A signal-transmission system.
$\dot{\theta}_2 = S_m\mu_m\theta_1/[(S/j\omega) + R + j\omega J]$
$\theta_2 = S_m\mu_m\theta_1/[S + pR + p^2J]$.

$$E = S_m\theta_1. \tag{1}$$

Now let us apply this electrical force to the armature of a direct-current motor. The motor now generates a torque (a mechanical force) that is proportional to the input voltage (an electrical force). We may treat the ratio of these two forces as we would a numeric, and call it the amplification factor of the motor.

$$T = \mu_m E. \tag{2}$$

The circuit is shown in Fig. 1.

Let us compare the potentiometer, the motor, and a conventional amplifying tube. They all require some bias or excitation: a power supply for the potentiometer, a field for the motor, and a plate battery for the tube. All three have input terminals, though the input terminal of the potentiometer is mechanical. All three have output terminals, though the output terminal of the motor is mechanical. All three generate a proportional force, but it does not necessarily appear at the output terminals. To measure it we must prevent motion by connecting an infinite impedance. This means open circuiting the potentiometer and the tube, and stalling the motor. With all three, the ratio of generated force to input tends to be constant with good design. All three have an internal resistance. It can be measured as the ratio of open-circuit voltage to short-circuit current, or stalled torque to free running speed.[1] In all three, the internal resistance tends to vary with amplitude.

Now let us connect the motor to a load. In this general discussion we will ignore reduction gears just as we would ignore an output transformer. The impedance of the load mesh may include a stiffness, such as a spring that restores the output shaft to zero. It will include a resistance that is the sum of the motor's internal resistance and the friction of the load. It will include an inertia that is the sum of the motor inertia and the load inertia. We can write the familiar differential equation relating displacement, impedance, and force

$$S\theta + R(d\theta/dt) + J(d^2\theta/dt^2) \doteq T. \tag{3}$$

Its solution is also familiar.[2]

$$\dot{\theta} = T/[(S/j\omega) + R + j\omega J]. \tag{4}$$

With the operational circuit notation we go from one of these equations to the other by the simple algebraic manipulation of using p for the derivative operator d/dt, using the same p for the reactance operator $j\omega$, and then interpreting the results in whichever way makes sense. This is fairly reasonable, even to the nonmathematical engineer, since d/dt represents (on an incremental basis) division by time, and $j\omega$ represents (with phase shift) multiplication by frequency which is the reciprocal of time.

We will rewrite (3) and (4)

$$S\theta + Rp\theta + Jp^2\theta = T \tag{3'}$$

$$p\theta = T/[(S/p) + R + pJ] \tag{4'}$$

$$\theta = T/[S + pR + p^2J]. \tag{5}$$

This appearance of the resistance in association with p, or $j\omega$, is really not surprising. Suppose we had neither stiffness nor inertia, and that we tested the system with a small alternating input. The output velocity would follow the input. At very low frequency it would run up, and then wipe out, a large displacement in each cycle. At higher and higher frequencies, this cyclic displacement would be less and less. Moreover, the displacement will lag 90 degrees behind the input, because it is a maximum at the instant when the input and the velocity have returned to zero. When we are discussing displacement, resistance introduces both phase shift and dependence on frequency.

[1] W. L. Everitt, "Communication Engineering," McGraw-Hill Book Company, New York 18, N. Y., 1937, p. 46.
[2] $\dot{\theta}$ is used to represent angular velocity and ω is reserved for $2\pi f$.

We have an output motion that does reproduce the input signal if a constant spring stiffness is the controlling term in the denominator of (5). But if ωR or $\omega^2 J$ becomes significant, we get what the radio engineer calls "frequency discrimination." If the circuit elements are not constant, we get what the radio engineer calls "distortion." There may be extraneous disturbances that create "noise." Because of the low frequencies involved, these unwanted components can be observed directly, and we are conscious of their instantaneous values. So we call them error.

One way to reduce noise and distortion, or error, is by negative feedback. Suppose we mount the potentiometer body on the output shaft, while leaving the arm on the input shaft, as shown in Fig. 2. The input to the po-

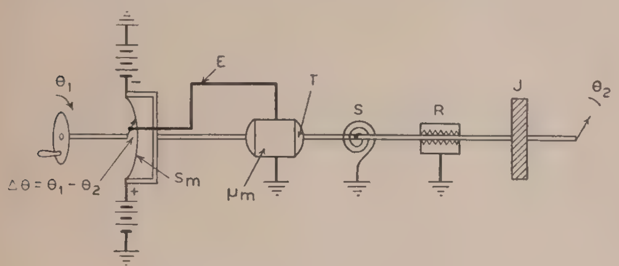


Fig. 2—A feedback system.
$\mu = -S_m\mu_m/[S + pR + p^2J]$
$\theta_2 = \mu(\theta_1 - \theta_2)$
$\Delta\theta = \theta/\mu.$

tentiometer is now the difference between the input and output displacements. This is usually called the error signal.

$$\Delta\theta = \theta_1 - \theta_2. \tag{6}$$

We shall arrange polarities so that the error signal, when transmitted through the potentiometer and motor, tends to reduce itself to zero.

This circuit is shown in Fig. 2. The potentiometer voltage is now dependent on both input and output displacements.

$$E = S_m(\theta_1 - \theta_2) \tag{7}$$

$$T = S_m\mu_m(\theta_1 - \theta_2) \tag{8}$$

$$\theta_2 = S_m\mu_m(\theta_1 - \theta_2)/[S + pR + p^2J]. \tag{9}$$

Let us introduce a new term, the loop amplification[3]

$$-\mu = -S_m\mu_m/[S + pR + p^2J]. \tag{10}$$

If we break the shaft between the motor and the potentiometer body, and displace the potentiometer end of the shaft some small amount, this signal will be transmitted around a loop through the potentiometer and motor and back to the break, where it will appear as a displacement just $-\mu$ times the displacement that was started. We could just as well break the wires between potentiometer and motor. A signal, now a voltage, started at the break would be transmitted around the loop and return as $-\mu$ times the voltage that was started.

[3] In the discussion of feedback it is customary to call the loop amplification $\mu\beta$. In this example, because of the direct feedback connection, $\beta = -1$. In many servo systems, β has other values.

Let us rewrite (6)

$$\Delta\theta = \theta/\mu. \tag{11}$$

This means that the error will be small if μ is large. If we make μ large enough we need not distinguish between input and output in computing error. For this reason the subscript on θ is omitted in (11). If we make μ large enough, we can tolerate, in the active amplifier, much of what we would call poor quality.

We can even omit the real stiffness on the output shaft, and greatly simplify the mechanical design. This makes $S=0$. In the remainder of this discussion we will consider S to be zero. This makes μ reactive, and the error will be out of phase with the signal. But if μ is large, the error will be small.

Let us evaluate this error more carefully. Let us rewrite (10)

$$\mu = \omega_0\omega_1/[p(p + \omega_1)] \tag{12}$$

where $\omega_0 = S_m\mu_m/R$, $\omega_1 = R/J$, and $S=0$. The value of μ is plotted in Fig. 3. If we put this value of μ in (11),

$$\Delta\theta = (p\theta/\omega_0) + [p^2\theta/(\omega_1\omega_0)]. \tag{13}$$

Equations (10) and (12) were in such a form that p could be easily interpreted as the reactance operator $j\omega$. Equation (13) is in such a form that p can be easily interpreted as the differential operator.

$$\Delta\theta = (\dot{\theta}/\omega_0) + [\ddot{\theta}/(\omega_1\omega_0)]. \tag{13'}$$

This shows the error to be composed of two parts: one proportional to velocity, and one proportional to acceleration. Both are dependent on the loop gain, and can be expressed in terms of two simple constants.

The loop gain has been plotted in Fig. 3. At low frequencies, where motion is limited by the resistance, velocity is proportional to error and independent of frequency, while displacement is inversely proportional to frequency. We have made the curve into a straight line by plotting the logarithm of amplitude ratio against the logarithm of frequency. Since doubling the frequency, which corresponds to an octave in music, results in halving the displacement, which is a 6-decibel loss to communication engineers, we say this low-frequency part of the curve has a slope of -6 decibels per octave. The intercept of this section of the curve is ω_0. This ω_0 is a useful measure of the low-frequency gain, and hence of the smallness of the error. In many designs, the inertia takes effect at some frequency less than ω_0, say at ω_1. Here the curve steepens to -12 decibels per octave, because the acceleration is proportional to the error, and hence the output displacement is inversely proportional to the square of frequency. If this section starts at ω_1, its intercept will be at the geometric mean of ω_0 and ω_1. Thus our errors depend on these intercepts of the loop gain curve; the larger the intercepts are, the smaller the errors will be.

The curve of Fig. 3 is, of course, an idealized curve.

In practice, there are still steeper sections at higher frequencies where other reactances take effect. Examples of such reactances are the inductance of the motor winding, and the shunt capacitances in vacuum-tube amplifiers that may be inserted between the potentiometer and the motor. The true curve is not a series of straight-line segments, but a smooth curve that rounds off the corners between them.

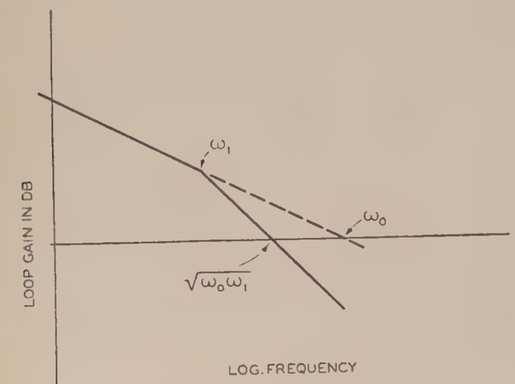


Fig. 3—Simple loop-gain characteristics.
$\mu = \omega_0\omega_1/[p(p+\omega_1)]$
$\Delta\theta = \theta/\mu = (p\theta/\omega_0) + (p^2\theta/\omega_0\omega_1)$.

We have examined error. Let us examine the other main element in servo design, stability. If we combine (9) and (10), we obtain

$$\begin{aligned} \theta_2 &= \mu(\theta_1 - \theta_2) \\ &= [\mu/(1+\mu)]\theta_1. \end{aligned} \tag{14}$$

We have been working with the idea that if μ is numerically large, then θ_2 and θ_1 are very much alike. But suppose the absolute value of μ is 1. This occurs at the intercept of the loop gain curve. The reactances that reduced the gain will also produce phase shift. Suppose that at the loop-gain intercept, the phase shift is just 180 degrees. Now μ has the value -1. In (14), θ_2 may now have a finite value, while θ_1 is zero; that is, at this intercept frequency we get an output with no input. The system oscillates. Because the frequency is so low, we sometimes say the system hunts. By analogy to an oscillating audio-frequency amplifier we may say it "sings."

To emphasize the mechanism of the oscillation, we say it sings around the loop. We designed the loop with a phase reversal at low frequencies. If we had an error, it was propagated around the loop and, because of this reversal, annulled itself. But at this singing frequency, the reactances have introduced an additional phase reversal, and when the error gets around the loop it tries to reinforce itself. If the return signal is larger than the original error, the error grows until it is limited by overloading in the system. If the return signal is weaker than the original error, the error will die out and disappear. The weaker the return signal, the faster the error will decay.[4]

This leads us to describe two margins of safety that we need.[5] At gain crossover, which is the frequency for which the amplification is unity and the gain is zero, we need a phase margin. And at phase crossover, which is the frequency for which the return signal is in phase with the error, we need a gain margin. It is good design practice to have a gain margin of 10 to 20 decibels and a phase margin of 40 to 60 degrees.

In Fig. 4, the curve marked "gain I" shows the loop gain for a typical servo system. In this case the potentiometer was excited with alternating current, and between the potentiometer and motor were a vacuum-tube amplifier, a detector, and a filter. The filter caused the additional steepening of the gain curve at high frequencies.

In Fig. 4, the curve marked "phase I" shows the loop-phase characteristic. It is of considerable importance that if the gain characteristic is known, the phase characteristic can be computed from it.[6] This means that if we can make amplitude measurement of the loop gain as a function of frequency, we can compute our phase margins. It means even more: if we can determine the loop gain at one frequency, and if we can determine the various "corner frequencies" at which the gain curve changes slope, then we can draw both curves completely and determine both gain and phase margins.

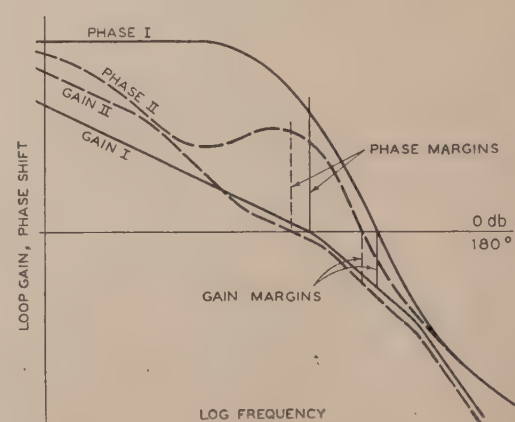


Fig. 4—Loop characteristics
I. before equalization
II. after equalization.

Bode gives general formulas relating gain, phase, and frequency. In most servo designs, it is easy to compute the characteristics of each independent part separately and then simply add them together; that is, it is easy if the circuit constants are known. Too often the servo designer attempts his design without knowledge of these circuit constants. Their determination should be a first step in the design. These circuit constants include mechanical inertia, mechanical resistance, the mutual or conversion stiffness of potentiometers, and the amplification or torque factor of motors, as well as the more familiar electrical constants.

[4] H. Nyquist, "Regeneration theory," *Bell Sys. Tech. Jour.*, vol. 11, p. 126; January, 1932.

[5] H. W. Bode, "Relations between attenuation and phase in feedback amplifier design," *Bell Sys. Tech. Jour.*, vol. 19, pp. 421–455; July, 1940.

[6] There exists a notable exception to this statement, but it seldom annoys the servo designer. See footnote reference (5).

In Fig. 4, curves I have been drawn with satisfactory margins but with low intercepts and hence large errors. To reduce errors at low frequencies we can increase the gain, for example, in the vacuum-tube amplifier. But this will destroy the margins and cause singing at some higher frequency.

Can we introduce a phase-correcting network that will improve the phase margin and permit the use of more gain? Yes, we can. And, by analogy to amplifier and wire transmission practice, we will call that network an equalizer. The basis of this equalization is as follows: Bode has shown that, with the exception already noted, phase shift is associated with the slope of the gain curve in the amount of 15 degrees of phase shift for each decibel per octave of slope. To have a 45-degree phase margin we must have, in addition to the initial reversal, a phase shift of 135 degrees. This must be associated with a slope of 9 decibels per octave.

In Fig. 4, curves II show the gain and phase of a servo loop after equalization. Here loss was introduced at a frequency just below the 6-to-12 corner of the original curve. This was done by means of a shunt capacitor in a direct-current part of the system. It gave a steeper slope and more phase shift. Then a resistance was put in series with this capacitor so that we would get a flat loss at higher frequencies and recover the lower phase shift.

It will be observed that in the region of gain cross-over we have a moderate slope in the gain curve and hence a safe phase margin. But at a little lower frequency, we have a steep section. This has raised the low-frequency end of the curve, thereby giving us a large low-frequency μ, and hence small errors.

A Very-High-Frequency Aircraft Antenna for the Reception of 109-Megacycle Localizer Signals*

BRUCE E. MONTGOMERY†, MEMBER, I.R.E.

***Summary*—A brief review of an instrument landing system is given to show where the localizer antenna fits into the over-all scheme. The localizer-antenna requirements are stated, and definitions that apply to this antenna are given. A description of an antenna that meets the specified requirements is given, including curves that show its performance over a ground plane. Patterns taken in flight on a twin-engine transport plane are included.**

Introduction

An INSTRUMENT landing system for aircraft may be made up of three parts: first, a localizer transmitter to provide horizontal guidance; second, a glide-path transmitter to provide vertical guidance; and third, marker transmitters to provide spot checks on the progress of the aircraft towards a successful landing. The localizer is placed a short distance off the far end of the airport runway on which the landing is to be made and projects its course down the center of this runway. The glide-path transmitter is located near the localizer and projects its sloping course in proper relation to the localizer course. The intersection of the planes of the localizer and glide-path courses is the line the aircraft follows to a landing.* Marker transmitters are placed at two points on the landing path. One of them is several miles from the edge of the airport and the other is at the airport edge. These transmitters project beams upward, and through a suitable receiver aboard the aircraft notify the pilot of his position.

This paper will confine its attention to the antenna mounted on the aircraft to receive the localizer signals. In the process of finding a suitable antenna, a number of types were investigated. Photographs and a description of the successful antenna developed and curves and data showing its performance are included.

The successful antenna must meet the following requirements:

1. The size, weight, and aerodynamic drag must be kept to a minimum.
2. A band of frequencies from 108.3 to 110.3 megacycles must be received.
3. The antenna must be sensitive to horizontally polarized waves only.
4. The output from the antenna delivered to the receiver must not be more than 10 decibels below the the standard dipole described later in this paper. This must hold throughout the band described in (2).
5. The pattern in the horizontal plane must be free from variations in output greater than 6 decibels for any heading in relation to the signal source. This requirement is to be met over a plane conducting surface. Maximum pickup should occur fore and aft.

Requirement (1) is obvious without discussion. Requirement (2) was determined by the frequency-band assignment made for localizer operation. Requirement (3) was fixed by the type of polarization used in the transmitting antenna. Previous investigation[1,2] had

* Decimal classification: R525. Original manuscript received by the Institute, March 22, 1945; revised manuscript received, June 8, 1945.

† United Air Lines, Chicago, Illinois.

[1] C. H. Jackson and J. M. Lee, "Preliminary investigation of the effects of wave polarization and site determination with the portable ultra-high-frequency radio range," *CAA Tech. Devel. Report 24*, February, 1940.

[2] P. C. Sandretto, "Principles of Aeronautical Engineering," McGraw-Hill Book Company, Inc., New York, N. Y., 1942, pp. 80–83.

shown horizontally polarized waves to be superior to those of vertical polarization. Requirement (4) was determined as follows: Flight tests were made with different antenna models. One of these models gave the minimum acceptable range for reception of localizer signals. The pickup of this antenna was then compared to the

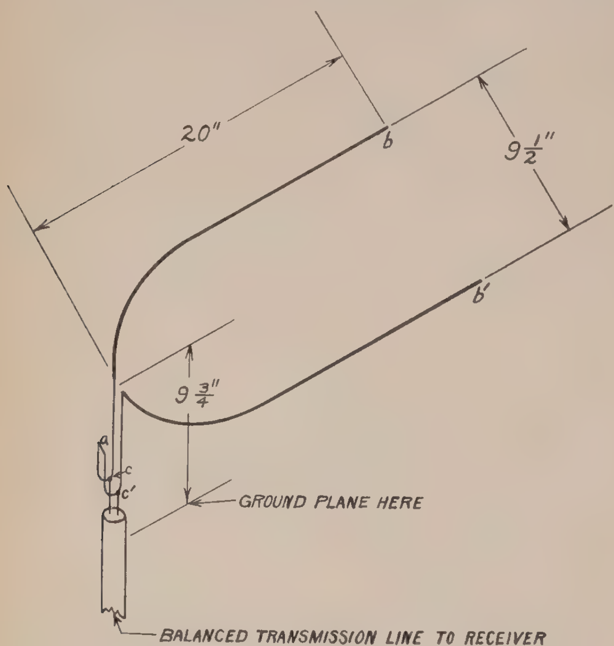


Fig. 1—Schematic diagram of the U antenna.

pickup of the standard dipole under standard test conditions and its output was found to be 10 decibels below the output of the standard dipole. Requirement (5) is necessary because the aircraft may be heading in any direction when first picking up the localizer course. The 6-decibel variation is an arbitrary figure arrived at after flight tests had shown that some variation could be accepted if it were not excessive.

Review of the Literature

Because of the relative newness of the very-high-frequency and instrument landing arts, there is little published information available on antennas of a type suitable for mounting on aircraft.

Alford and Kandoian[3] describe several types of loop antennas suitable for aircraft use. Their discussion is mainly limited to a mathematical derivation of the electrical properties, however.

Bennett[4] describes a similar loop antenna actually used on an aircraft in an instrument landing system. It is a horizontally mounted antenna, bent in the shape of a circle, and connected to a balanced transmission line in much the same manner as are the Alford loops.

Definitions

The *gain* of an antenna is defined as the output in decibels delivered at the antenna mid-band frequency using the output of the standard dipole as a zero-decibel reference. All gain measurements are made with the front of the antenna under test pointing towards the signal source.

The *selectivity* of an antenna is the output in decibels delivered over a frequency band using the output of the standard dipole at mid-band frequency as a zero-decibel reference. All selectivity measurements are made with the front of the antenna under test pointing towards the signal source. The signal-source-antenna current (at a current maximum) is held constant at all frequencies in the band.

The *horizontal space pattern* of an antenna is the plot of the output in decibels delivered at mid-band as the antenna is rotated through 360 degrees. The front of the antenna is taken as the 0-degree heading and the output at 0 degrees is used as the zero-decibel reference output. The space pattern may be taken over a plane conducting surface and will be referred to as the *free space pattern*, or it may be taken on an airplane and here it will be referred to as the *airplane space pattern*.

The U Antenna

This antenna meets the requirements set up in the preceding paragraphs. It is so named because its active pickup elements are in the shape of a U. It is mounted horizontally with the base of the U forward. It is shown schematically in Fig. 1. The antenna may be regarded as a quarter-wave resonant line that has been partially opened, or it may be regarded as a half-wave antenna partially folded. A current maximum occurs at a, and voltage maxima at bb'. There are points cc' across which the antenna resistance is equal to the characteristic impedance of the transmission line to the receiver. Fig. 2 shows an experimental model with the

Fig. 2—A laboratory model of the U antenna.

supporting mast removed to show the construction of the folded center section of the antenna. The height of the mast is $9\frac{3}{4}$ inches and the length of the U is 20 inches.

The U antenna in Fig. 2 was adjusted to have an input resistance of 230, 500, and 800 ohms (at point cc'

[3] Andrew Alford and A. G. Kandoian, "Ultra-high-frequency loop antennas," *A.I.E.E. Tech. Paper*, January, 1940.
[4] Robert P. Bennett, "Radio Signaling System," United States Patent No. 2,221,939, November 19, 1940.

in Fig. 1) and a selectivity curve was made for each condition. A balanced line of 200-ohm characteristic impedance was connected between the antenna and a source of radio-frequency energy. Standing waves were measured on this line and adjustments were made on the antenna to produce the required impedance. For instance, a standing-wave ratio of 2.5 with a voltage minimum a quarter wave from point cc' indicates a 500-ohm antenna resistance.

The selectivity curves were made as follows: A resistor whose impedance was measured as 201+j15 ohms at 110 megacycles was connected in place of the source of radio-frequency energy. Radiations were received on the antenna and the voltage developed across the load resistor was measured at several frequencies. These curves are shown in Fig. 3. The important points here are that as the input resistance of the antenna increases

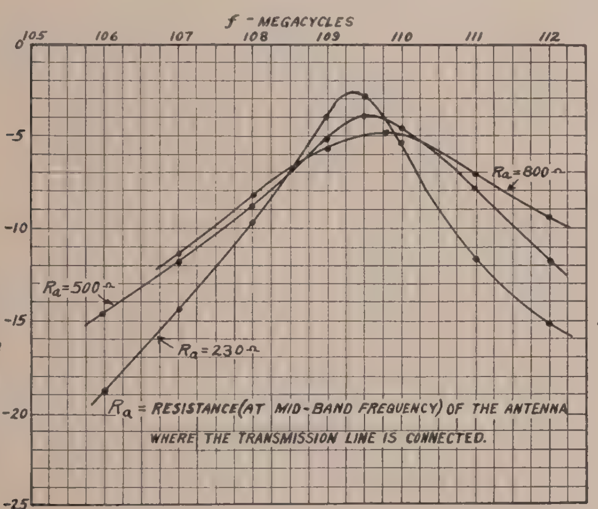


Fig. 3—Selectivity curves of the U antenna.

the output decreases at resonance, and the curve broadens. If fairly constant output is desired over a band several megacycles wide it is desirable to increase the antenna input resistance when the load, as represented by the receiver input, closely matches the line. If the load does not match the line, this method of broadening the antenna selectivity curve is not recommended, since the line will present a complex impedance to the antenna that will be a function of the length of the line. This may produce an undesirable shift in the selectivity curve of the antenna.

The zero-decibel output level in Fig. 3 is the output delivered to the load resistance when a horizontal half-wave dipole 18.5 inches high is substituted for the U antenna. This is the standard dipole to which previous reference has been made.

The free space pattern of the U antenna is shown in Fig. 4. It was made by rotating the antenna about a vertical axis while receiving energy in a field of constant intensity.

An improved laboratory model of the U antenna is shown in Fig. 5. The pickup arms are covered by polystyrene tubing, and the fore part of the antenna is covered by a housing formed from polystyrene sheet. The assembly is mounted on a streamlined mast.

Since this antenna must, when mounted on an airplane, receive signals under all types of flying conditions, a spray test was conducted to determine the ef-

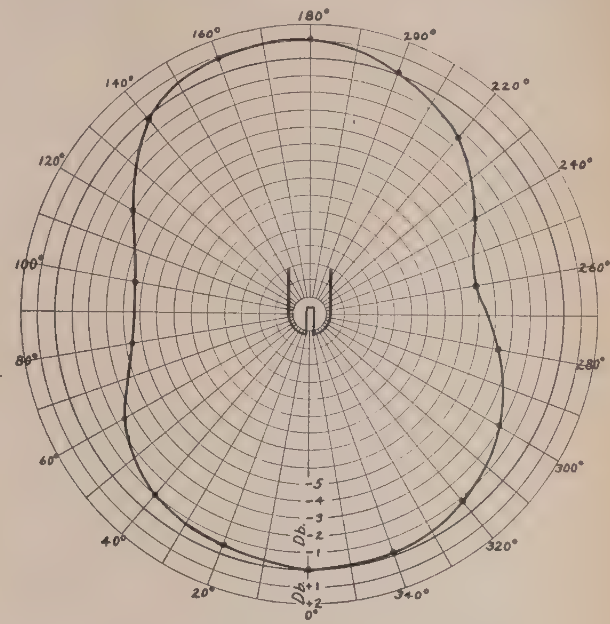


Fig. 4—Free space pattern of the U antenna.

fect of wetting the antenna while receiving signals. The antenna was mounted in position and transmitted energy received on it. The output when dry, wet under spray, wet—no spray, arms only wiped dry, and the

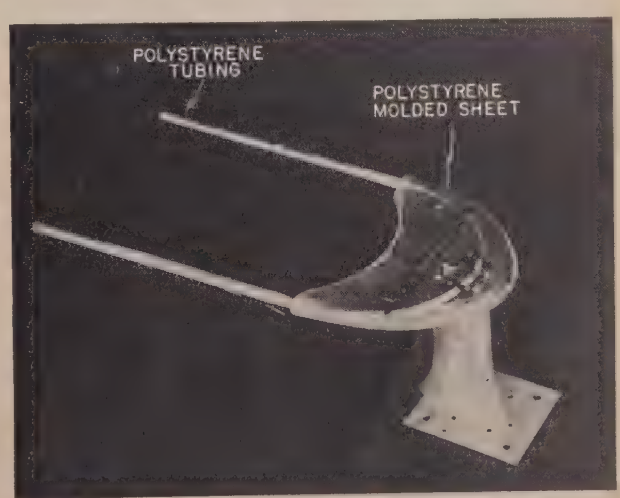


Fig. 5—An improved laboratory model of the U antenna.

entire antenna wiped dry was recorded. Arm coverings of one-half and three-quarter inch diameter polystyrene and one-inch diameter bakelite tubing were used. Table I shows the result of these tests.

The following conclusions can be drawn from these data: most of the decrease in output is caused by water

TABLE I

Type of Arm Covers	Entire Antenna Wet Under Spray	Entire Antenna Wet—No Spray	Arms Only Wiped Dry	Entire Antenna Wiped Dry
	Decibels	Decibels	Decibels	Decibels
No Cover	−4.3	−2.2	—	−0.5
1/2-inch diameter polystyrene	−3.4	−1.7	−0.3	0
3/4-inch diameter polystyrene	−3.5	−2.9	−0.7	−0.35
1-inch diameter bakelite	−2.3	−0.8	—	0

on the arms; the water on the housing has little effect; the largest diameter covering gives the best results.

One determination made that is not shown in Table I is the loss introduced by the one-inch-diameter bakelite tubes. The bakelite covers caused a 5-decibel decrease in output as compared to no covers, while the polystyrene covers caused little or no decrease in the output.

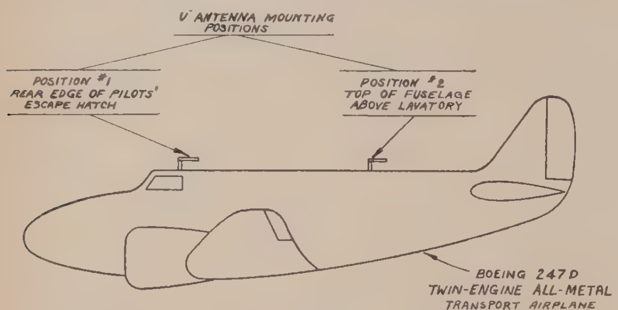


Fig. 6—Antenna mounting positions for Figs. 7, 8, 9, and 10.

Fig. 8—Airplane space pattern of the U antenna obtained when flying a tight circle with the antenna mounted in position 1, Fig. 6.

This information indicates that an arm covering of relatively large diameter is desirable, if loss in output due to water is to be avoided, and that polystyrene is a satisfactory material, while bakelite is not.

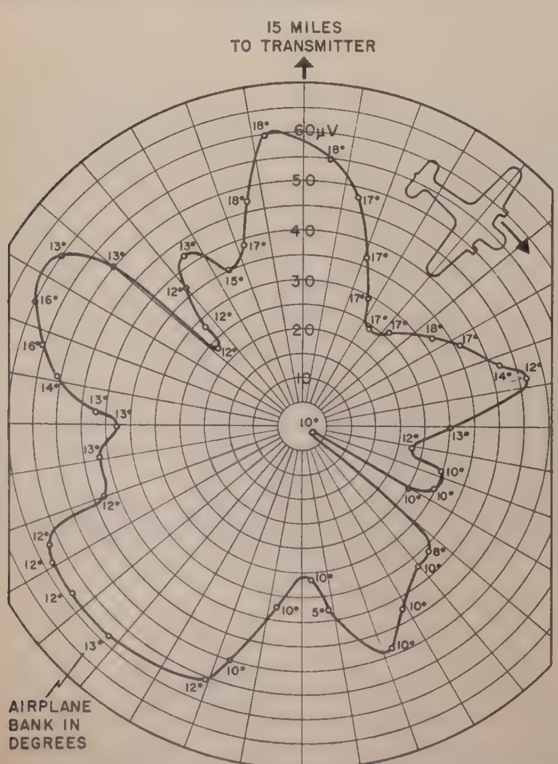


Fig. 7—Airplane space pattern of the U antenna obtained when flying a flat circle with the antenna mounted in position 1, Fig. 6.

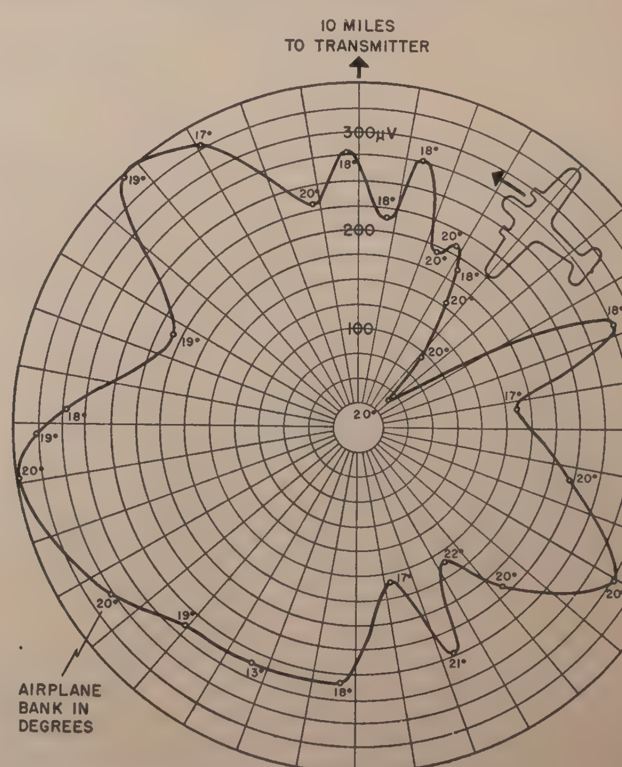


Fig. 9—Airplane space pattern of the U antenna obtained when flying a flat circle with the antenna mounted in position 2, Fig. 6.

It was also found that all types of arm covers lowered the resonant frequency of the antenna by approximately 3 megacycles. This means that removal (in service) of the arm covers, due to breakage, will cause a serious detuning of the antenna that may result in 8 or 9 decibels loss.

The airplane space pattern of the U antenna is of interest. Patterns on an all-metal twin-engined Boeing

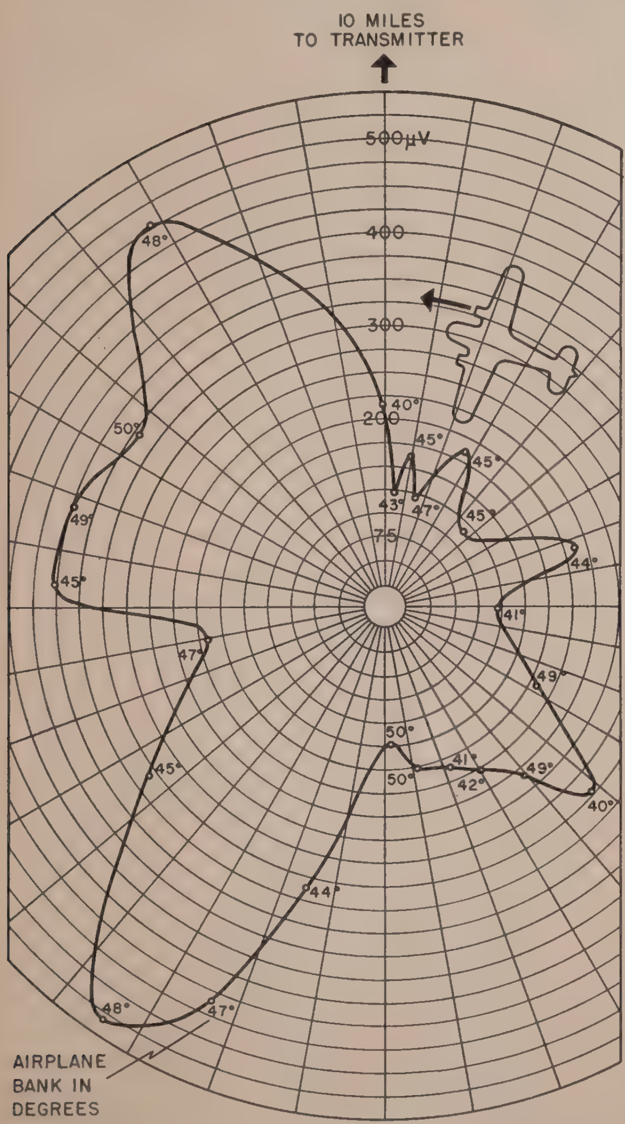


Fig. 10—Airplane space pattern of the U antenna obtained when flying a tight circle with the antenna mounted in position 2, Fig. 6.

247D airplane are included for two mounting positions: the rear edge of the pilots' escape hatch and the top of the fuselage above the lavatory. Fig. 6 shows these two mounting positions. Flight tests were made which resulted in Figs. 7, 8, 9, and 10. To obtain data for these, circles were flown at 10 to 15 miles from the transmitter operating on 109.5 megacycles. The direction to the transmitter is shown by the arrow on each figure. On each antenna two circles were flown. One was a flat circle (the airplane was generally banked less than 20 degrees) and the other was a tight circle (the airplane was generally banked more than 40 degrees).

These patterns are interpreted as follows: the strength of the signal delivered to the receiver is given by the length of the radial line from the center of the graph to any point on the curve. The flight path of the airplane is indicated by the airplane outline and arrow on each illustration: For instance, in Fig. 7, the strength of the signal received, when the airplane is in the position shown, is proportional to the length of the line from the center of the graph to the curve measured along the radial passing through the wings of the airplane. The numbers appearing near the curve indicate the angle of bank at that point in the circle. When the airplane was in the position shown, the angle of bank is seen to be 17 degrees. The direction of arrival of the received signal at the airplane is 230 degrees, where 0 degrees is the direction in which the airplane is heading.

Fig. 11—The military version of the U antenna.

Figs. 7, 8, 9, and 10 are representative of the patterns that are obtained from this type of antenna when mounted on an airplane. Considerable irregularity is observed. The pattern in three dimensions is quite irregular since the shape of the pattern is shown to be influenced considerably by the angle of bank in which the airplane is flying. No attempt has been made to predict the pattern that any particular mounting position might give. Even after obtaining a pattern, it is sometimes found that the pattern is contrary to what is expected. For example, in Fig. 7, more signal is obtained when the airplane is flying across course with the left wing raised in the transmission path than when the airplane is flying in the opposite direction with the wing lowered out of the transmission path. It is probable that the lowered wing, which is large in relation to a wavelength, is reflecting some energy into the antenna, and that the path length is such as to cause partial cancellation when combined with the direct wave at the antenna. Generally

speaking, the antenna should be mounted on or near the center line of the airplane. Better results are usually obtained with the antenna mounted on top of the fuselage, although satisfactory results may be obtained with mounting on the underside of the fuselage if nonradio factors require it.

Fig. 11 shows the AN-100-A antenna. This U antenna has the pickup arms supported in a molded-rubber head which does introduce some loss but provides good mechanical support. The aerodynamic drag produced by this antenna is about 5 pounds at 200 miles per hour. This will cause the speed of a transport plane of the DC-3 type to be reduced less than one-half mile per hour. This antenna was produced in quantity by Communication Equipment and Engineering Company, of Chicago, Illinois, for the Army Air Forces.

ACKNOWLEDGMENTS

The author wishes to thank the director of the Aircraft Radio Laboratory at Wright Field, Dayton, Ohio, for allowing the publication of this material.

The patterns given in Figs. 7, 8, 9, and 10 were taken under the direction of Lieutenant-Colonel F. L. Moseley by radio engineers of the Aircraft Radio Laboratory, while the antenna was mounted on the United Air Lines' flight-research airplane.

A Proposed Standard Dummy Antenna for Testing Aircraft-Radio Transmitters*

CHANDLER STEWART, JR.†, ASSOCIATE, I.R.E.

Summary—**A new type of dummy antenna employing a 35-foot roll of coaxial cable and a power indicator is described. It roughly simulates the impedance characteristics of an actual aircraft antenna, which is used over the range of 2 to 30 megacycles. Its impedance characteristics are not appreciably affected by mechanical shock, humidity, ageing, etc. Power measurements can be made with it over a wide impedance range with a single indicating instrument. The impedance presented to the transmitter terminals is unaffected by lead geometry, ammeter impedances, etc. Its power-dissipating capability is limited by the flow temperature of the cable dielectric, and is of the order of 125 watts.**

INTRODUCTION

THE RESISTANCE of a fixed-wire aircraft antenna over the range of the usual communication frequencies from about 2 to 20 megacycles is quite likely to vary as much as from 1 to 10,000 ohms, and the reactance from −5000 to +5000 ohms.[1,2] (See Fig. 1). Even greater impedance variations than these are common with certain types of fixed aircraft antennas. Consequently, the output-impedance tuning and loading range requirements of aircraft-radio transmitters which are designed to operate with any of these antennas must meet especially stringent requirements. Since it is inconvenient to test transmitters by connecting them to typical aircraft antennas and determining whether they will deliver the required power output over the required frequency range, and since such a test method would cause serious and unlawful interference with actual communications, it has been common practice to use dummy antennas for this purpose.[1–3] Such dummy antennas have consisted of networks of lumped resistance and reactance, whose complexity depended upon the bandwidths and the extent to which the dummies simulated real antennas.

Testing of aircraft-radio transmitters with dummy

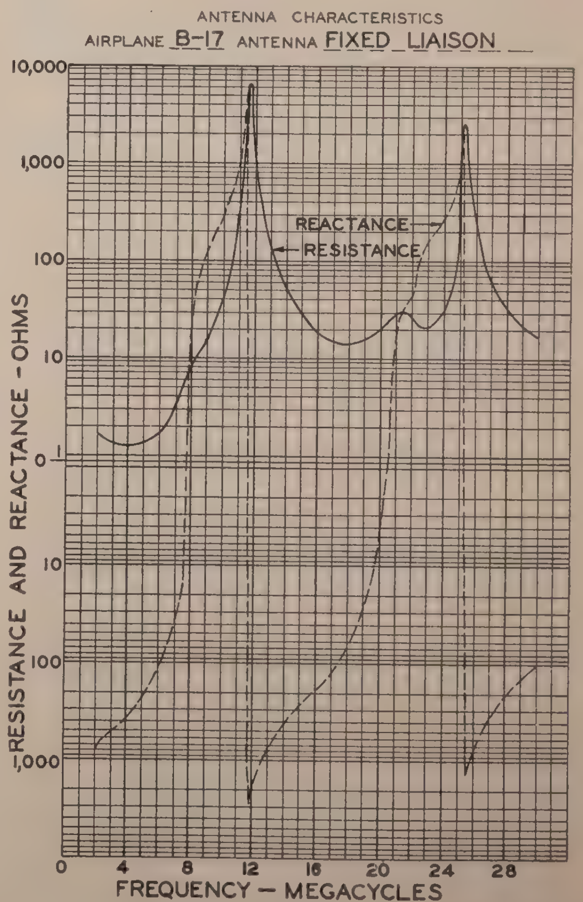


Fig. 1—The characteristics of impedance versus frequency for a typical aircraft-communication antenna.

* Decimal classification: R525×R327. Original manuscript received by the Institute, April 25, 1945; revised manuscript received, June 20, 1945.

† Headquarters, Air Technical Service Command, Wright Field, Dayton, Ohio.

[1] P. J. Holmes, "Aircraft antennas characteristics," *Electronics*, vol. 15, pp. 46–48; December, 1942.

[2] G. L. Haller, "Aircraft antennas," Proc. I.R.E., vol. 30, pp. 357–362; August, 1942.

[3] H. Salinger, "A dummy-dipole network," Proc. I.R.E., vol. 32, pp. 115–116; February, 1944.

antennas of this type has been subject to the following limitations:

(1) Simple two- or three-element networks are not capable of simulating the half-wave and odd quarter-wave impedances of a real antenna at both high- and low-frequency regions of the communication band of 2 to 20 megacycles.

(2) The more complex networks must be large physically, because of the high voltages each element must stand.

(3) The complex networks are difficult to construct in a way that will insure close adherence to a standard impedance characteristic under conditions of normal use. This is an especial problem at integral half-wave-length (maximum-impedance) regions, where a discrepancy in the reactance of a capacitor or inductor of only one per cent (due either to manufacturing tolerances or to subsequent shifts in physical characteristics due to normal handling) could easily double or triple the terminal impedance of the dummy.

(4) Due to the extremely wide resistance ranges involved, power measurement over the required frequency range has necessitated a set of three or four ammeters, and has required great care to avoid damaging the lower-range meters by overload.

(5) Because the impedance "seen" by the transmitter depends upon the geometry of the connecting leads and the ammeter impedance, as well as the terminal impedance of the dummy antenna, tests on exactly similar transmitters with the same type of dummy have frequently yielded widely differing results.

In order to reduce these limitations and to simplify the problem of standardization on a single type of dummy antenna for testing aircraft-radio transmitters operating in the range of from 2 to 30 megacycles, a new type of dummy antenna is proposed.

Although this proposed dummy antenna has fixed characteristics which represent only one type of aircraft antenna, its maximum and minimum resistance and reactance values approach the limits encountered in nearly all aircraft antenna in its frequency range. Therefore, all types of aircraft transmitters for these frequencies must be capable of operating satisfactorily at these impedance limits. For many of the same reasons that a standard dummy antenna has been established for testing receivers,[4] a standard dummy antenna for aircraft-radio transmitters seems desirable.

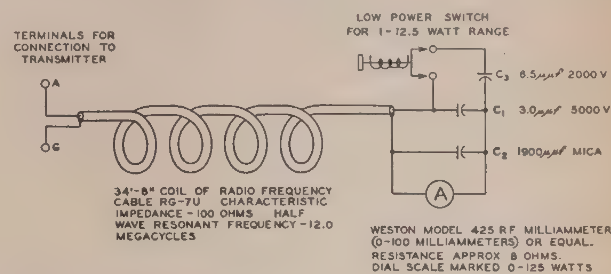
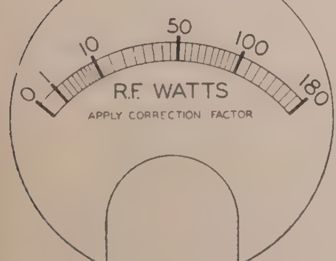


Fig. 2—Schematic diagram of the proposed dummy antenna, employing a thermomilliammeter power indicator.

[4] I.R.E. "Standards on Radio Receivers," 1938.



Special power scale for diode indicator.

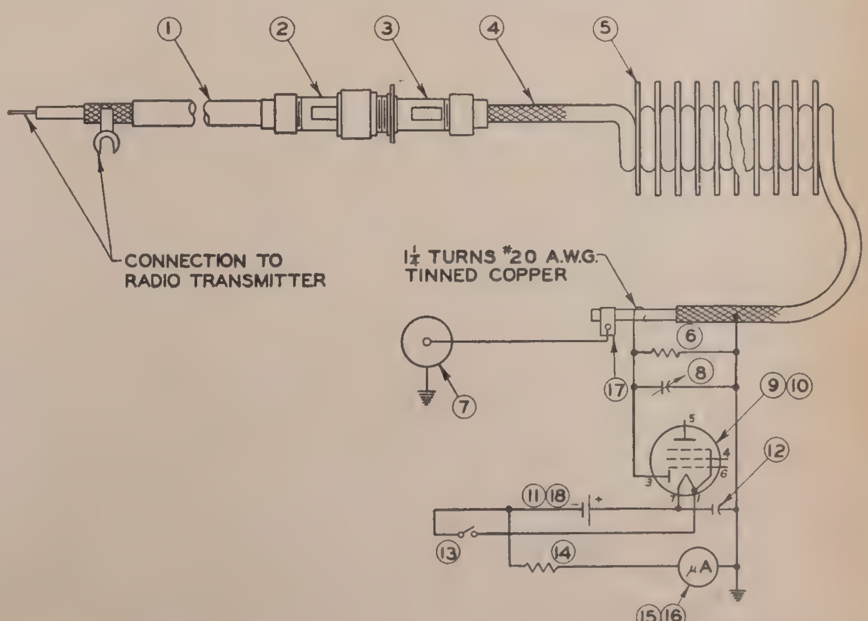


Fig. 3—Schematic diagram of the dummy antenna, with a diode power indicator.

1. Approximately 3½ feet of radio-frequency cable RG-7/U for connection to the transmitter under test.
2. High-voltage coaxial connector.
3. High-voltage coaxial connector.
4. Approximately 31 feet of radio-frequency cable RG-7/U with the outer insulating jacket removed. The total length of the cable is such as to resonate at 12 megacycles.
5. Copper heat-radiating fins.
6. Resistance R_1, approximately 4000 ohms.
7. Connection for oscilloscope or modulation monitor.
8. Adjustable capacitance, approximately ½ to 1 micromicrofarad.
9. Radio tube, type 1S5.
11. No. 6 dry cell.
12. By-pass capacitor, 500 micromicrofarads.
13. Filament switch.
14. Voltmeter multiplier.
15. Microammeter, with special power scale.

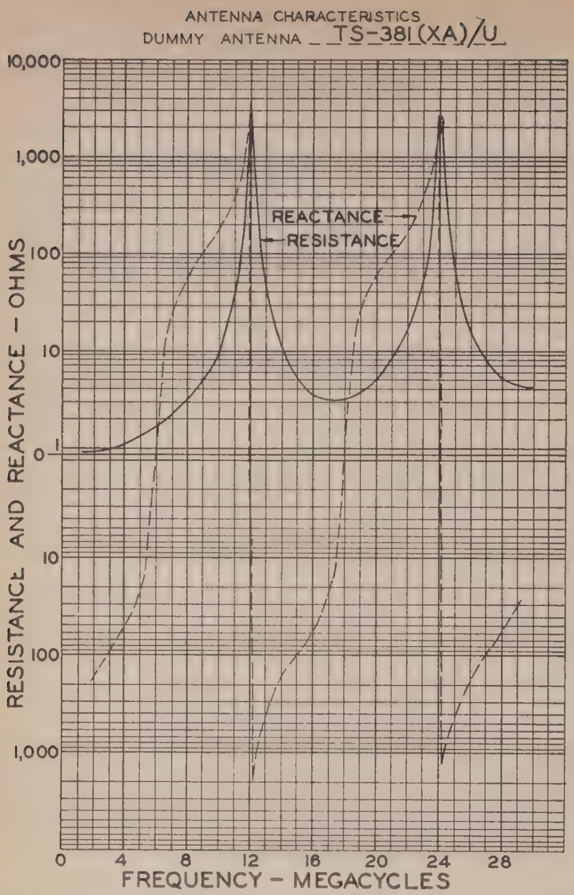


Fig. 4—The characteristics of impedance versus frequency for the proposed dummy antenna.

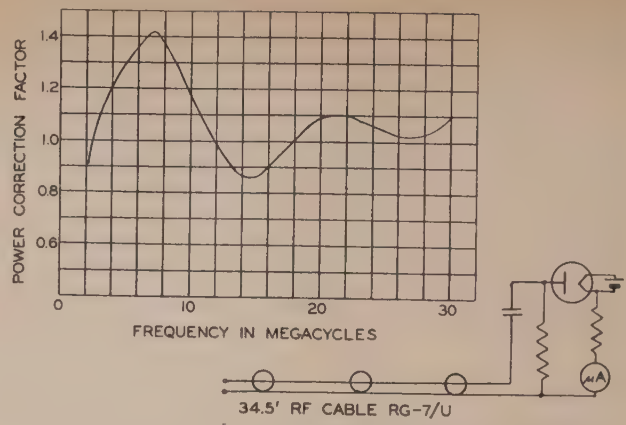


Fig. 6—Correction factor to be applied to the power indications of the diode power indicator of Fig. 3.

Description

This dummy antenna, which is designed to simulate roughly a typical 40-foot fixed aircraft antenna, consists of approximately 34.5 feet of radio-frequency cable RG-7/U. Most of this length is stripped of its outer insulating jacket, and wound around a form with metal heat-radiating fins between turns. One end of this cable is connected to the transmitter under test, and the other through a small series capacitor to a power indicator, such as a shunted radio-frequency milliammeter or a diode voltmeter, as shown in the alternative circuit arrangements of Figs. 2 and 3, respectively.

The impedance characteristic is shown in Fig. 4. Note the similarity between the impedance characteristics of this dummy and of the typical aircraft antenna shown

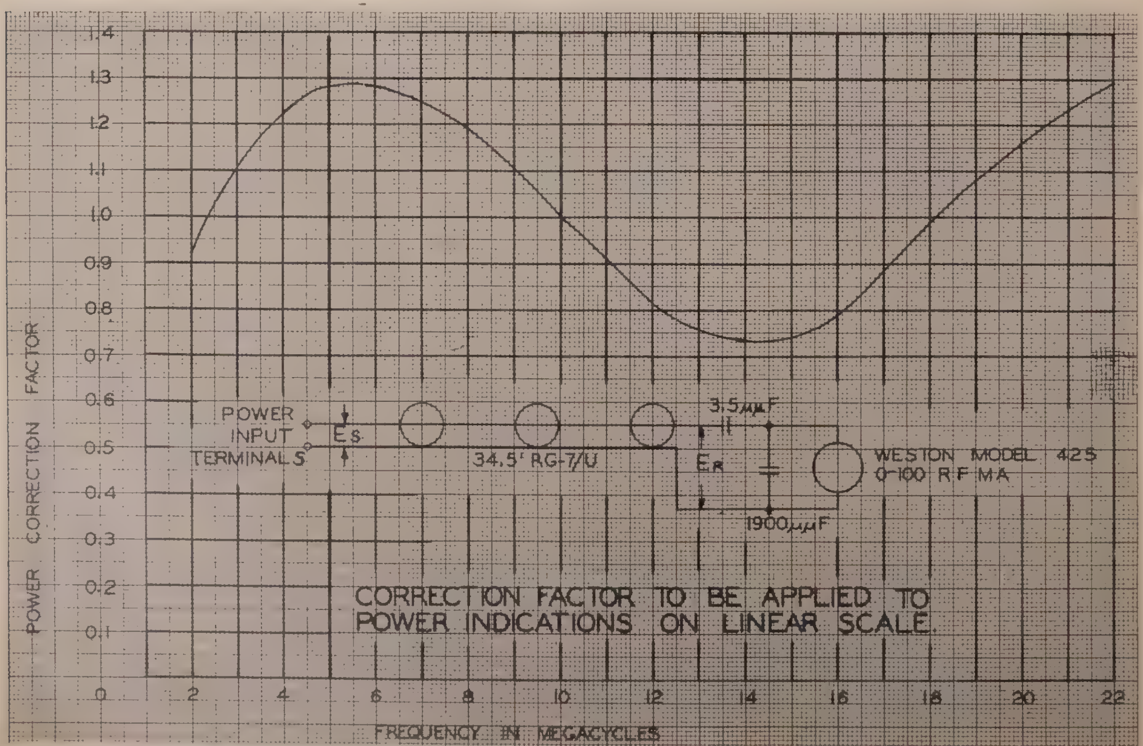


Fig. 5—Correction factor to be applied to the power indications of the thermomilliammeter power indicator of Fig. 2.

in Fig. 1. A correction, such as shown in Figs. 5 and 6, must be applied to the power indications.

Mathematical Theory of the Dummy Antenna

1. Definitions of Terms

α = attenuation constant of cable in nepers per foot $\cong 0.00115N$

β = wavelength constant of cable in radians per foot

C = distributed capacitance of cable in farads per foot

C_1 = power-indicator series capacitance in farads

C_{pk} = total diode shunt capacitance

$\gamma = \alpha + j\beta$

E_1 = radio-frequency voltage input to diode, root-mean-square

E_r = radio-frequency voltage at open end of cable, root-mean-square

f = frequency of applied voltage in cycles per second

G = distributed conductance of cable in mhos per foot

I_r = current at open end of cable $= 0$

I_s = current input to cable in amperes

$j = \sqrt{-1}$

l = total length of cable in feet

n = a whole integer, such that $\beta l - n\pi = 0$

N = attenuation constant of cable in decibels per 100 feet

P = power input to cable in watts

P_f = power factor of cable dielectric

R = cable-conductor resistance in ohms per foot

R_1 = effective diode input resistance in ohms

R_{max} = maximum input resistance of cable in ohms

R_{oc} = input resistance of cable in ohms

ω = angular velocity of input voltage in radians per second $= 2\pi f$

X_{max} = maximum input reactance of cable in ohms

X_{oc} = input reactance of cable in ohms

Z_o = complex characteristic impedance of cable in ohms

$|Z_o|$ = absolute magnitude of cable characteristic impedance, in ohms

Z_{oc} = input impedance of cable in ohms $= R_{oc} + jX_{oc}$

2. Derivation of Cable Open-Circuit Impedance

From basic transmission-line theory,[5] the following group of formulas may be obtained:

$$\alpha = (1/2)(G\sqrt{L/C} + R\sqrt{C/L}), \tag{1}$$

$$\beta = \omega\sqrt{LC}, \tag{2}$$

and

$$Z_o = \sqrt{L/C}\,\underline{/(1/2)[(G/\omega C) - (R/\omega L)]}. \tag{3}$$

From basic capacitor theory,[6]

$$G = P_f \omega C, \tag{4}$$

and

$$P_f = G/\omega C. \tag{5}$$

Dividing (1) by (2),

$$\alpha/\beta = (1/2)[(G/\omega C) + (R/\omega L)]. \tag{6}$$

[5] W. L. Everitt, "Communication Engineering," McGraw-Hill Book Company, New York 18, N. Y., 1937, p. 118.

[6] F. E. Terman, "Radio Engineers' Handbook," McGraw-Hill Book Company, New York 18, N. Y., 1943, p. 110.

Subtracting (6) from (5) yields

$$P_f - (\alpha/\beta) = (1/2)[(G/\omega C) - (R/\omega L)]. \tag{7}$$

Combining (3) and (7) produces

$$Z_o = \sqrt{L/C}\,\underline{/P_f - (\alpha/\beta)}, \tag{8}$$

which, by definition, becomes

$$Z_o = |Z_o|\,\underline{/P_f - (\alpha/\beta)}. \tag{9}$$

In practice it will be found that

$$P_f \ll 1 \tag{10}$$

and

$$\alpha/\beta \ll 1, \tag{11}$$

so that the following approximation of (9) is accurate:

$$Z_o \cong |Z_o| \{1 + j[P_f - (\alpha/\beta)]\}. \tag{12}$$

Combining (2) and (8), and applying (10) and (11) yields

$$\beta l = \omega C Z_o l. \tag{13}$$

From transmission-line theory[7]

$$Z_{oc} = Z_o \coth \gamma l. \tag{14}$$

Substituting (12) in (14) yields, by definition of γ,

$$Z_{oc} \cong Z_o\left[1 + j\left(P_f - \frac{\alpha}{\beta}\right)\right]\frac{(\sinh 2\alpha l - j \sin 2\beta l)}{2(\sinh^2 \alpha l + \sin^2 \beta l)}. \tag{15}$$

The cable used will be short enough for

$$\alpha l \ll 1 \tag{16}$$

and since (10) and (11) also apply, the following approximations are justified, from (15):

$$R_{oc} = \frac{Z_o[2\alpha l + (P_f - \alpha/\beta)\sin 2\beta l]}{2[(\alpha l)^2 + \sin^2 \beta l]} \tag{17}$$

and

$$X_{oc} = (-Z_o \sin 2\beta l)/\{2[(\alpha l)^2 + \sin^2 \beta l]\}. \tag{18}$$

3. Regions Far from Half-Wave Points

For frequencies such that

$$\sin \beta l \gg \alpha l, \tag{19}$$

(17) becomes

$$R_{oc} \cong Z_o\{(\alpha l/\sin^2 \beta l) + [P_f - (\alpha/\beta)] \cot \beta l\} \tag{20}$$

and (18) becomes

$$X_{oc} \cong -Z_o \cot \beta l. \tag{21}$$

4. Regions Near Integral Half-Wave Points

For frequencies such that

$$\beta l - n\pi \cong 0 \tag{22}$$

where

$$n \geq 1, \tag{23}$$

$$\sin \beta l = \beta l - n\pi, \tag{24}$$

(17) becomes

$$R_{oc} \cong \{Z_o[P_f(\beta l - n\pi) + (\alpha/\beta)n\pi]\}/[(\alpha l)^2 + (\beta l - n\pi)^2] \tag{25}$$

and (18) becomes

$$X_{oc} \cong \{-Z_o(\beta l - n\pi)\}/\{(\alpha l)^2 + (\beta l - n\pi)^2\}. \tag{26}$$

The values of βl for maximum resistance are obtained by equating the first derivative of (25) to zero. In this

[7] See p. 159, equation (46), of footnote reference 5.

case, the assumptions of (10), (11), and (22) permit the following approximation:

$$(dR_{oc})/(d\beta l) \cong \{2Z_o \alpha l n\pi[(n\pi/\beta l)-1]\}/[(\alpha l)^2+(\beta l-n\pi)^2]^2=0, \quad (27)$$

from which

$$\beta l = n\pi. \quad (28)$$

The value of the maximum resistance is obtained by substituting (28) in (25)

$$R_{\max} = Z_o/\alpha l. \quad (29)$$

The values of βl for maximum reactance are obtained by equating the first derivative of (26) to zero

$$(dX_{oc})/(d\beta l) = Z_0[(\beta l-n\pi)^2-(\alpha l)^2]/[(\alpha l)^2+(\beta l-n\pi)^2]^2=0, \quad (30)$$

from which

$$\beta l - n\pi = \pm \alpha l. \quad (31)$$

The value of the maximum reactance is obtained by substituting (31) in (26)

$$X_{\max} \cong \pm Z_o/(2\alpha l). \quad (32)$$

It is interesting to note that, from (29) and (32),

$$R_{\max} \cong 2X_{\max}. \quad (33)$$

5. Power Input to Open-Circuited Cable

From transmission-line theory,[8]

$$I_s = I_r \cosh \gamma l + (E_r/Z_o) \sinh \gamma l. \quad (34)$$

Since, for an open-circuited cable, $I_r=0$, this becomes, by definition of γ,

$$I_s = (E_r/Z_o)(\sinh \alpha l \cos \beta l + j \cosh \alpha l \sin \beta l), \quad (35)$$

from which, applying (16),

$$|I_s|^2 = (E_r^2/Z_o^2)[(\alpha l)^2 + \sin^2 \beta l]. \quad (36)$$

$$P = |I_s|^2 R_{oc}. \quad (37)$$

Substituting (17) and (36) in (37) yields

$$P = (E_r^2/Z_o)\{\alpha l + [P_f - (\alpha/\beta)][(\sin 2\beta l)/2]\}. \quad (38)$$

6. Calibration of Diode Indicator

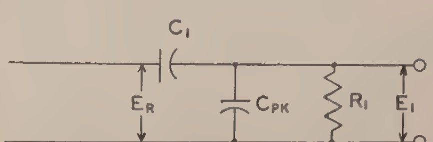


Fig. 7—Equivalent radio-frequency circuit of the diode voltmeter.

Application of basic theory to the diode-voltmeter circuit, shown in Fig. 7, yields

$$|E_r/E_1| = |1 + (C_{pk}/C_1) - (j/\omega C_1 R_1)| \quad (39)$$

which, substituted in (38), produces

$$P = E_1^2/Z_o[\alpha l - (P_f - \alpha/\beta)(\sin 2\beta l)/2] \cdot (1 + C_{pk}/C_1)^2[1 + 1/[\omega(C_1 + C_{pk})R_1]^2]. \quad (40)$$

Experimental Procedure and Results

The impedance characteristics of this dummy antenna, as shown in Fig. 4, were calculated from formulas (20), (21), (25), and (26), and verified by direct measurements with General Radio radio-frequency bridges type 821 and type 916, and with Boonton Q-Meter type 160A. The curve of cable open-end voltage of Fig. 8 was obtained from (38) and verified experimentally.

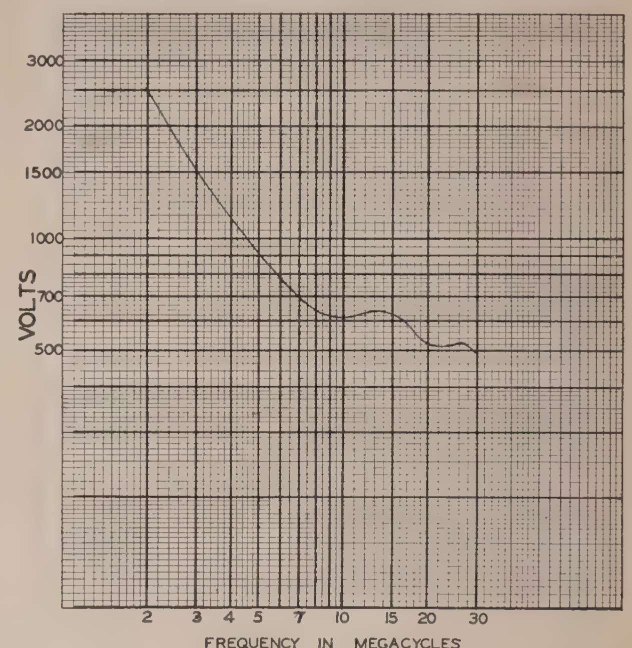


Fig. 8—Open-end voltage of the 34.5-foot length of radio-frequency cable RG-7/U with 100 watts input.

It is apparent that, since the power into the dummy is proportional to the square of the open-end voltage, and since the deflection of a thermocouple-type milliammeter is proportional to the square of its current, insertion of such an instrument into a linear network receiving voltage from the cable open end would result in a deflection proportional to power. However, since this deflection is also a function of frequency, as shown in Fig. 8, a correction must be applied to the meter deflection. In order to minimize the magnitude of this correction, a search was made for a coupling network which would have a transfer impedance versus frequency characteristic similar to the curve of Fig. 8. Several combinations of shunt and series capacitors were tried, and their response determined experimentally. The results of tests on some of these are shown in Figs. 9 and 10. It was found that the inductive reactances of the type of small mica shunt capacitors used were of the same order of magnitude as their projected capacitive reactances and the meter resistances, and that these reactances were considerably affected by the geometry of the connections to the meter terminals, the lead lengths, etc. Consequently, it was neither possible to measure the capacitor characteristics nor to calculate response curves such as those of Figs. 9 and 10. Since the 1900-micromicrofarad curve of Fig. 9 appeared most nearly to match the curve of Fig. 8, this capacitance was used in an experimental model of the dummy, although thermomilliammeters of different manufacture, shunted by other values of capacitors, may yield curves equally

[8] See p. 159, equation (46), of footnote reference 5.

as suitable. The correction-factor curve of Fig. 5 was obtained from these data for this arrangement, which used a special linear scale which was attached to the radio-frequency milliammeter.

An alternative power indicator is shown schematically in Fig. 3. It employs a vacuum-tube voltmeter, whose

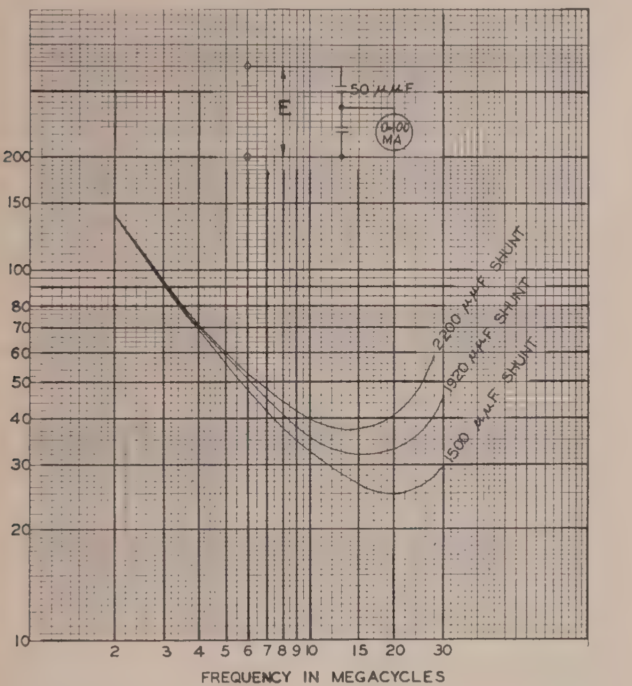


Fig. 9—Calibration of Weston model 425 0-125 radio-frequency milliammeter No. 128502 in a capacitive network. Direct-current resistance = 8.2 ohms.

frequency response is given by (39). The power calibration of this indicator versus frequency, in contrast to that of the previously described thermomilliammeter circuit, can be calculated, and (40) can be used for this purpose. Since the deflections of this meter are proportional to voltage, the dial scale must have power indications proportional to the square of the deflection. This permits a considerably greater power range on one scale than does the thermomilliammeter arrangement.

A third possible power-indicating arrangement, consisting of a thermomilliammeter capacitively coupled to the cable at a point about one fourth the way from its open end, would be simplest to design since lead inductances or meter resistances would not affect its calibration. However, this was not the subject of any experimentation at this time, because of the limited range of frequencies over which satisfactory indications could be obtained.

Power-dissipation tests were made on this dummy, employing twenty copper fins, each eight inches square. At an ambient of 71 degrees centigrade, with 125 watts input, a maximum rise of 18 degrees centigrade was encountered, which is well within the safe limits for polyethylene. The latter does not distort appreciably at 100 degrees centigrade.

Complete data on the effect of temperature on the input impedance has not been collected as yet. One manufacturer has stated that the cable characteristic impedance can be held to within two per cent of a nominal value of 100 ohms, so that, from (17) and (18), it would seem that the input impedance of the dummy antenna could also be held to within close limits.

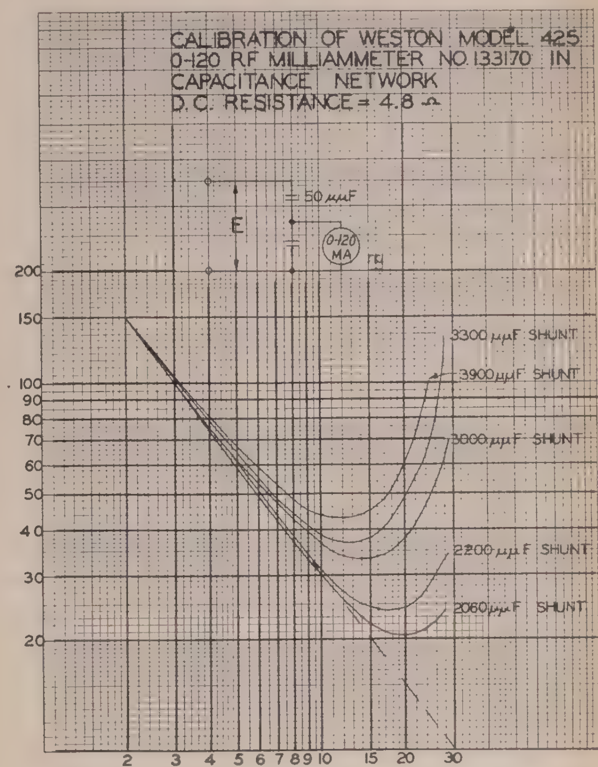


Fig. 10—Calibration of a radio-frequency milliammeter in a capacitive network.

Conclusions

This proposed standard dummy antenna appears to have the following advantages over other types used in testing aircraft-radio transmitters:

(1) It more nearly simulates the impedance characteristics of an actual aircraft antenna over a wide frequency range than does any other simple device yet proposed.

(2) The impedance characteristics are not appreciably affected by mechanical shock, humidity, aging, etc.

(3) Power measurements over extremely wide impedance ranges can be obtained with a single indicating instrument.

(4) Since the cable connects directly to the terminals of the transmitter under test, without any opportunity for lead geometry variation and without any input ammeters, the impedance "seen" by the transmitter is essentially that given by a curve such as that of Fig. 4.

For general use in the field, the diode indicator is probably preferable to the thermomilliammeter type, since it permits a considerably greater power range, and replacement meters and capacitors need not meet as strict impedance requirements, permitting the use of standard components.

Some Considerations Concerning the Internal Impedance of the Cathode Follower*

HAROLD GOLDBERG†, SENIOR MEMBER, I.R.E.

Summary—The behavior of the cathode follower working into a load consisting of R and C in parallel is investigated for step-function and sine-wave input. It is found that the tube is easily cut off for applied voltages which are decreasing functions of time. The conditions for which the tube is conductive for decreasing applied voltages are derived. They are severe for step-function input but not for sine-wave input. The influence of tube parameters, supply voltages, applied-voltage wave form and amplitude, and load conditions are analyzed. Design recommendations are suggested which increase the conductive range. It is pointed out that these recommendations are of value even if conditions for conduction during all operating situations cannot be satisfied.

IT IS WELL known that the internal, or source impedance presented by the cathode follower is $R_p/(\mu+1)$ or $1/(G_m+G_p)$. Since G_m is large in comparison with G_p, the expression is usually given as $1/G_m$. While it is obvious that the tube can present such a source impedance only if it is conducting, it is an error to assume that the follower is conducting as long as proper direct operating voltages are applied and the grid-to-ground, or driving voltage, does not become more negative than the grid cutoff voltage of the tube. This error may be commonly made because the steady-state direct voltage necessary to cut the follower off is the grid cutoff voltage of the tube.

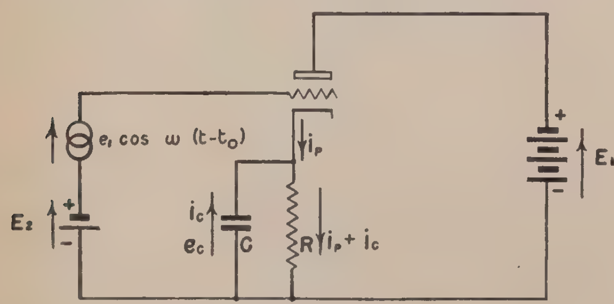


Fig. 1—General circuit.

Experience with the behavior of the cathode follower driving a load consisting of R and C in shunt and driven by a rectangular wave soon shows that care must be taken to prevent cutoff during the times the driving voltage is decreasing. One sees evidence of tube cutoff even though the driving voltage to ground is far from cutoff. This may also happen with a sinusoidal driving voltage, but not to such a great degree. The reason, of course, is that the cathode-to-grid voltage of the follower may exceed cutoff even though the driving voltage does not, and this behavior is a function not only of the operating voltages applied to the tube, but of the wave form and amplitude of the driving voltage, the characteristics of the load, and the tube parameters.

It is the purpose of this paper to analyze the conditions under which cutoff takes place and to point out what factors govern this action. An ideal triode has been assumed, that is, one whose current equation may be written $i_p=(\mu E_g+E_p)G_p$ as long as i_p is nonnegative. It may be said that such a choice is highly idealized, and that the nonlinear behavior in the region of cutoff should be taken into account, since this paper is concerned with the effects of cutoff. It turns out, however, that the choice of the ideal equation and its sharp cutoff characteristic predicts results which are qualitatively checked by experiment even if the quantitative results may be in error because of this idealization. Furthermore, the idealization leads to linear differential equations which may be solved, while the nonidealization leads to nonlinear differential equations whose solution may be very difficult. The final justification of the assumption is that it leads to information which is of value just as the ideal triode equation has led to valuable results many times, where it was not strictly justified.

The choice of load, R and C in parallel, may seem somewhat restricted too, but it covers a large number of applications of the follower. It serves to point out the dangers of too indiscriminate use of the follower and shows that the follower is not the panacea for all ills, as it sometimes seems. This paper is not intended as a comprehensive treatment of the action of cathode followers, but is presented as an analysis of one aspect of the operation of these devices.

The analysis will be concerned with an ideal triode connected as a cathode follower to a cathode load consisting of R and C in shunt. A direct plate-supply voltage is provided and a driving voltage is superimposed on a direct grid-supply voltage to ground. The operation is analyzed for a step voltage which instantaneously raises the grid-to-ground voltage, a step voltage which instantaneously lowers the grid-to-ground voltage, and a voltage which is a sinusoidal function of time. The first two driving voltages are of interest because of their applications to rectangular waves and pulses. The sinusoidal driving voltage is of traditional interest and also throws light to the operation of the so-called "infinite-impedance" detector.

The discussion will be based on the following derivations and set of definitions (see Fig. 1):

E_g = the instantaneous voltage between cathode and grid

E_p = the instantaneous voltage between cathode and plate

* Decimal classification: R132. Original manuscript received by the Institute, November 9, 1944; revised manuscript received, January 30, 1945. Presented, National Electronics Conference, Chicago, Ill., October 6, 1944. (The Chicago Section of The Institute of Radio Engineers was one of the sponsors of the National Electronics Conference.)

† Bendix Radio Division, Bendix Aviation Corporation, Baltimore, Md. Formerly, Stromberg-Carlson Company, Rochester, N. Y.

R_p = the dynamic plate resistance
$G_p=1/R_p$
G_m = the grid-plate transconductance
μ = the amplification factor
$G_m=\mu G_p$
$e=2.7183$, the base of natural logarithms
i_p = the instantaneous cathode current in the arrow direction
i_c = the instantaneous capacitor current in the arrow direction
e_c = the instantaneous capacitor voltage in the arrow direction
E_b = the steady plate-supply voltage in the arrow direction
E_2 = the steady input voltage in the arrow direction
e_1 = the amplitude of the sinusoidal input voltage in the arrow direction

It will be convenient to define four more quantities which occur frequently in the analysis.

$E^*=\mu E_2+E_b$
$E_1^*=\mu E_2-\mu\Delta E_2+E_b$
$R_e=RR_p/[R_p+R(\mu+1)]$ R_e is the parallel combination of R and the internal impedance of the cathode follower $R_p/(\mu+1)$
$S=\omega CR_e$

The basic equation for the ideal triode is

$$i_p=(\mu E_g+E_p)G_p. \tag{1}$$

Since the tube is a nonlinear device, all subsequent equations involving the tube must be restricted as follows: The plate current of the tube in the arrow direction is identically zero unless $i_p\geqq 0$; $E_p\geqq 0$. These conditions state that the effective plate voltage is never negative and that the plate or cathode current is never negative. In every case, effects due to transit time will be neglected.

The following circuit equations may now be written:

$$i_p=G_p[-\mu e_c+\mu E_2+\mu e_1\cos\omega(t-t_0)-e_c+E_b] \tag{2}$$

$$e_c-(i_p+i_c)R=0 \tag{3}$$

$$i_c=-Cde_c/dt. \tag{4}$$

Solutions for e_c and i_p may be written in the form

$$e_c=\alpha_1 e^{-g(t-t_0)}+\beta_1\cos[\omega(t-t_0)-\theta_1]+\delta_1 \tag{5}$$

$$i_p=\alpha_2 e^{-g(t-t_0)}+\beta_2\cos[\omega(t-t_0)+\theta_2]+\delta_2. \tag{6}$$

Substituting (5) and (6) in (2), (3), and (4) and equating the coefficients of like functions of t gives

$$\delta_1=R_eE^*/R_p \tag{7}$$

$$\delta_2=R_eE^*/RR_p \tag{8}$$

$$\alpha_2=-(\mu+1)\alpha_1/R_p \tag{9}$$

$$g=1/CR_e \tag{10}$$

$$\beta_1=\mu e_1R_e/R_p\sqrt{1+S^2} \tag{11}$$

$$\beta_2=\mu e_1R_e\sqrt{1+S^2R^2/R_e^2}/[RR_p\sqrt{1+S^2}] \tag{12}$$

$$\theta_1=\tan^{-1}S \tag{13}$$

$$\theta_2=\tan^{-1}S(R/R_e-1)/(1+S^2R/R_e). \tag{14}$$

This gives

$$e_c=\alpha_1\epsilon^{-(t-t_0)/CR_e}+(\mu e_1R_e/R_p\sqrt{1+S^2})\cos[\omega(t-t_0)-\theta_1] +R_eE^*/R_p \tag{15}$$

$$i_p=[-(\mu+1)\alpha_1/R_p]e^{-(t-t_0)/CR_e} +\{(\mu e_1R_e/[RR_p\sqrt{1+S^2}])\sqrt{1+S^2R^2/R_e^2}\} \cdot\cos[\omega(t-t_0)+\theta_2]+R_eE^*/RR_p. \tag{16}$$

These solutions hold for all t's greater than or equal to t_0. α_1 may be determined from the initial conditions at t_0. At t_0, the capacitor voltage is assumed to be e_{c0}. α_1 is then obtained from

$$\alpha_1=e_{c0}-R_eE^*/R_p-\mu e_1R_e\cos\theta_1/(R_p\sqrt{1+S^2}) \tag{17}$$

since $\cos\theta_1=1/\sqrt{1+S^2}$

$$\alpha_1=e_{c0}-R_eE^*/R_p-\mu e_1R_e/[R_p(1+S^2)]. \tag{18}$$

These equations assume that the tube is conducting; i.e., that i_p is not negative. If the tube is not conducting, the equations become

$$e_c=e_{cT}e^{-(t-T)/CR} \quad \text{for} \quad T\leqq t \tag{19}$$

where e_{cT} is the capacitor voltage at time T.

Case 1: The investigation will be for a step-function voltage or voltages applied to the grid, plate, or combination (see E^*). While it may seem that this case is somewhat trivial, it is applicable to certain types of sweep circuits and serves to point out certain features of the operation of the cathode follower. The internal impedance of the follower is also a direct consequence of the treatment of this case. (See Fig. 2.)

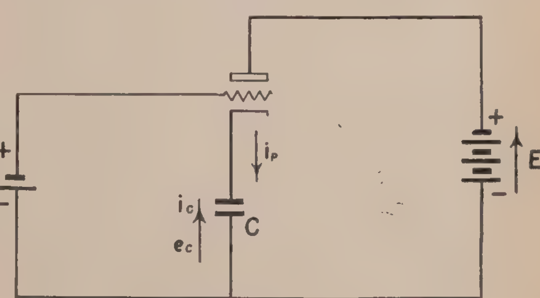


Fig. 2—Circuit, case 1.

For this case we have

$$R_e=R_p/(\mu+1)=1/(G_m+G_p) \tag{20}$$

$$g=(\mu+1)/CR_p \tag{21}$$

$$e_c=[e_{c0}-E^*/(\mu+1)]e^{-(\mu+1)(t-t_0)/CR_p}+E^*/(\mu+1) \tag{22}$$

$$i_p=\{-(\mu+1)[e_{c0}-E^*/(\mu+1)]/R_p\}e^{-(\mu+1)(t-t_0)/CR_p} \tag{23}$$

$$=-i_c.$$

It is at once evident that unless e_{c0} is less than $E^*/(\mu+1)$, nothing happens when E^* is applied to the tube. If e_{c0} is equal to or greater than $E^*/(\mu+1)$, the current must be zero or negative. For either of these cases, the current is zero. This states that the capacitor may be charged under certain conditions to a voltage higher than its initial voltage, but may never be discharged. It is evident that discharge would require the transfer of electrons from plate to cathode. The charge is exponential with a time constant $CR_p/(\mu+1)$. From the analogy of a capacitor charging through a resistance R, which gives rise to a time constant CR, the tube may be seen to act as a resistance $R_p/(\mu+1)$ or $1/(G_m+G_p)$. This may be called the internal impedance of the cathode follower. The effective voltage of the

follower as a generator is given by $E^*/(\mu+1)$. It is evident that the tube conducts only when the applied step function results in voltages on the tube which allow conduction. If any charging of the capacitor takes place, it will be because the tube conducts, and the charge will be an exponential function of time. The equilibrium, or steady state, is reached when the tube ceases to conduct.

This entire behavior, except for the time constant of the charge, may be predicted without deriving the equations. The current in a capacitor charged through a resistor is a unidirectional exponential pulse which gradually approaches zero as the capacitor voltage approaches its final value. Furthermore, a capacitor cannot charge to a voltage greater than that which will render the tube nonconducting in the circuit of Fig. 2. These two facts demand that the capacitor charge exponentially to a final value which just requires zero current in the tube.

Case 2: In this case, R is finite. The investigation is again for a voltage which is applied stepwise in time to the circuit. If E^* is applied to the circuit at t_0, and e_{c0} is the capacitor voltage at this instant, the circuit equations are, for $t \geqq t_0$

$$e_c = (e_{c0} - R_eE^*/R_p)e^{-(t-t_0)/CR_e} + R_eE^*/R_p \tag{24}$$

$$i_p = \{[-(\mu+1)(e_{c0} - R_eE^*/R_p)]/R_p\}e^{-(t-t_0)/CR_e} + R_eE^*/RR_p. \tag{25}$$

The charge is again seen to be exponential in nature. The time constant is CR_e. In this instance, the internal impedance of the follower in parallel with R forms the resistive part of the time constant. If e_{c0} is less than R_eE^*/R_p, the capacitor charges upward to the final value R_eE^*/R_p and i_p is never negative. It is evident from (25), however, that e_{c0} may also be greater than R_eE^*/R_p without requiring i_p to be negative. *This means that C may also be discharged under certain conditions without cutting the tube off. The use of a finite resistance in shunt with C allows discharge of C under conditions in which the tube conducts and presents a low impedance in shunt with C, just as it did in the charging case. This means that the application of decreasing step voltages to the follower, subject to restrictions, will not cut the follower off if C is shunted by a finite resistance R.* The conditions for which this is true may be investigated with the aid of the following equations. Let us suppose that E^* had been applied at some time prior to t_o and the circuit allowed to come to equilibrium. The capacitor voltage would be R_eE^*/R_p. Now at t_0, change the value of E^* stepwise downward to a new value E_1^*. This might be done by changing E_2 or E_b or both. The greatest interest is in the case corresponding to a change in E_2 alone. The equations for this case are

$$e_c = (ReE^*/R_p - R_eE_1^*/R_p)e^{-(t-t_0)/CR_e} + R_eE_1^*/R_p \tag{26}$$

$$i_p = [-(\mu+1)(R_eE^*/R_p - R_eE_1^*/R_p)/R_p]e^{-(t-t_0)/CR_e} + R_eE_1^*/RR_p. \tag{27}$$

The limits for the applied decreasing step function for which the tube will still conduct may now be determined by setting i_p at $t=t_0$ equal to zero and solving for ΔE^* which is $E^*-E_1^*$. This gives the maximum value of ΔE^*, Δ_0E^*, that may be used without stopping conduction in the tube. We obtain

$$\Delta_0E^* = E_1^*R_p/(\mu+1)R. \tag{28}$$

However, this gives

$$\Delta_0E^* = R_eE^*/R \tag{29}$$

but

$$e_{c0} = R_eE^*/R_p$$

therefore

$$\Delta_0E^* = R_pe_{c0}/R. \tag{30}$$

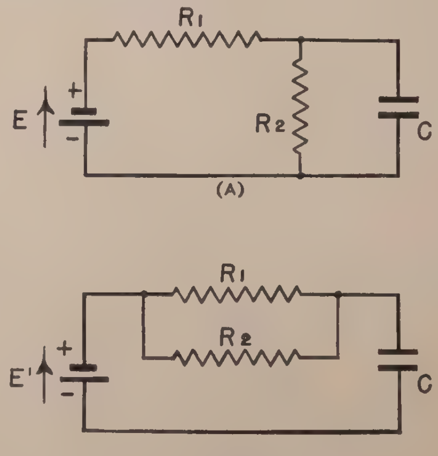
In particular, the interesting case is for Δ_0E^*, E_b fixed. For this case, Δ_0E^* becomes $\mu\Delta_0E_2$ and we obtain

$$\Delta_0E_2 = R_pe_{c0}/\mu R \tag{31}$$

$$\Delta_0E_2 = e_{c0}/RG_m. \tag{32}$$

This states that the conducting range is increased by increasing E^* and by decreasing R and G_m. It is unfortunate that the latter is true, since a decrease in G_m results in an increase in the time constant of the charge or discharge. To insure conduction, G_m may not be chosen as large as possible for low impedance without taking into account its effect on Δ_0E_2. The smallest permissible value of R may determine the type of tube to be used. Unless R is reduced to such a value that RG_m is equal to unity, Δ_0E_2 is not very large. When RG_m is equal to unity, Δ_0E_2 is that change in E_2 which causes the capacitor to discharge to zero.

The question may be asked as to the physical mechanism which gives rise to this action. It is evident that the discharge current cannot flow through the tube. Yet, the capacitor discharges as though the discharge current does flow through the tube. This may be explained by considering the action of another circuit first. Consider the circuits in Figs. 3(A) and 3(B).



$E' = \dfrac{R_2 E}{R_1 + R_2}$

(B)

Fig. 3.

It is well known that, if C is charged by a two-terminal black box containing the network to the left of C in either circuit, the effect is exactly the same. In the case of Fig. 3(A), however, all of the charging current must pass through R_1 and yet the time constant is given by the product of C and the parallel combination of R_1 and R_2. In this case the applied voltage is higher than for the case of 3(B). It is enough higher so that the initial

charging current is the same for both cases. In other words, the capacitor of 4(A) charges toward a voltage E with a time constant of CR_1 initially, but it need charge only to a final voltage of E'. The capacitor voltage will, therefore, reach $E'(1-1/e)$ in a shorter time than it would take to reach $E(1-1/e)$ if R_2 were not in the circuit. In essence, the shorter time constant is the consequence of an apparent equilibrium condition at the commencement of charging which is greater in value than the actual equilibrium condition. A similar action takes place in the discharge condition in the follower. The initial apparent equilibrium value is lower than the actual, and the capacitor approaches its actual equilibrium value with greater speed than expected. The condition circuitwise is equivalent to the tube conducting in the reverse direction.

There are applications, however, where it is not possible to meet the condition for tube conduction. Such a situation is one in which the applied ΔE^* is greater than $\Delta_0 E^*$. The tube ceases to conduct for a time in this case, until the conditions for conduction are again restored by the fall of the capacitor voltage. Let us suppose that ΔE^* is applied for a time T_1-t_0 following t_0 and is then removed, and let us suppose that not only is ΔE^* greater than $\Delta_0 E^*$ but is such that for the entire interval T_1-t_0,

$$-(\mu+1)e_c + E_1^* < 0. \quad (33)$$

Then i_p is identically zero over the interval and the capacitor voltage is given by

$$e_c = (R_e E^*/R_p)e^{-(t-t_0)/CR}. \quad (34)$$

The rapidity of the discharge may be increased only by decreasing the value of R for a fixed C. The internal impedance of the tube is effectively infinite in this case.

Suppose, however, that ΔE^* has such a value that at a time t_0+T, e_c has fallen to such a value that for all subsequent times in the interval (T, T_1)

$$-(\mu+1)e_c + E_1^* > 0. \quad (35)$$

Then for the interval T_1-T, the tube conducts and the equations for the interval T_1-T are

$$e_c = (e_{cT} - R_e E_1^*/R_p)e^{-(t-T)/CR_e} + R_e E_1^*/R_p \quad (36)$$

$$i_p = [-(\mu+1)(e_{cT} - R_e E_1^*/R_p)/R_p]e^{-(t-T)/CR_e} + R_e E_1^*/RR_p \quad (37)$$

where $$-(\mu+1)e_{cT} + E_1^* = 0. \quad (38)$$

The equilibrium value is not zero in (36) *as it was in the case of* (34).

The interval $T-t_0$ may be calculated from (34) and (38). It is

$$T - t_0 = CR \log_e [R_e(\mu+1)E^*/R_p E_1^*]. \quad (39)$$

If again the change from E^* to E_1^* is brought about by a change of E_2, ΔE_2, E_b fixed, (39) becomes

$$T - t_0 = CR \log_e [R_e(\mu+1)/R_p(1 - \mu\Delta E_2/E^*)]. \quad (40)$$

This equation tells us that for a fixed ΔE_2, $T-t_0$ may be made small by increasing E^* without limit. In fact, $T-t_0$ may be made to approach zero by increasing E^*. Equation (40) is not in such form as to make this evident, however. It has been previously postulated that ΔE_2 is greater than or equal to $\Delta_0 E_2$, then

$$\Delta E_2 \geqq R_e E^*/\mu R.$$

If this is true, then $\mu\Delta E_2/E^* \geqq R_e/R$

or $$\mu\Delta E_2/E^* = R_e/R + \eta$$

where η is some number greater than or equal to zero. If this is substituted in (40), and certain reductions carried out, we obtain

$$T - t_0 = CR \log_e [1/\{1 - \eta[1 + R_p/R(\mu+1)]\}]. \quad (41)$$

It is evident that the argument of $\log_e$ is always greater than or equal to unity and $T-t_0$ is therefore always greater than or equal to zero. $T-t_0$ will be zero when

$$\Delta E_2 = \Delta_0 E_2 = R_e E^*/\mu R$$

but this is the criterion for conductive operation and for this case $T-t_0$ must be zero.

Except for the case where the tube is nonconducting over the entire interval, T_1-t_0, the operation may be termed quasi conductive since the final phase of the discharge takes place with the tube conducting. As already pointed out, an increase of E^* decreases the time $T-t_0$, and also increases the conducting range. Therefore, even where the action is not entirely in the conducting range, the time of discharge may be reduced by increasing E^*. As the oscillograms show in Fig. 4, it is possible to start with the completely nonconducting case, and by increasing E^*, successively to shorten the time of discharge until conductive operation is attained. It is advantageous, therefore, in all cases, to keep E^* at its highest practical limit.

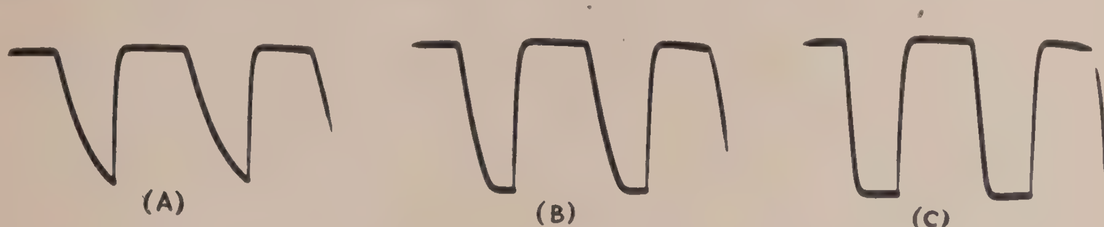


Fig. 4—Results of increasing E^* for rectangular wave input. (E^* increasing to right.)

Case 3: This case concerns itself with the application of sinusoidal input voltages. The analysis will again be directed toward the calculation of the region of conductive operation. This may be determined by the conditions which insure a nonnegative cathode current in the tube. An examination of (15) and (16) would seem to indicate that all of the terms in the cathode-current expression must be taken into account in determining this region. The exponential term may be ignored, however, and only the sinusoidal and steady terms considered.

This is based on the following argument. Consider e_1 to be zero at first and E^* to be applied and the circuit allowed to come to rest. Then increase e_1 at a very slow rate. If the increase is slow enough, we may intuitively see that a condition may be approached where the output voltage consists only of a sinusoidal and steady term and the tube is conducting at all times. In the limit, the condition is reached where the sum of the sinusoidal and steady terms are at all times greater than or equal to zero. This limit gives us the region of conductive operation. Any increase over this will clearly result in nonconduction over part of the cycle and the appearance of exponential terms in the current and voltage. The application of this argument gives us the criterion for conductive operation

$$(\mu R_e e_1/RR_p\sqrt{1+S^2})\sqrt{1+S^2R^2/R_e^2} \leqq R_e E^*/RR_p \quad (42)$$

or

$$e_1 \leqq E^*\sqrt{1+S^2}/[\mu\sqrt{1+S^2R^2/R_e^2}]. \quad (43)$$

When S is very small; i.e., when the applied frequency is very low, this becomes approximately

$$e_1 \leqq E^*/\mu \quad (44)$$

which states that the operation is equivalent to the conductive region for steady applied voltages. For steady voltages, an input voltage $-e_1$, applied in addition to E^*, of absolute value equal to E^*/μ, will cut the tube off. When S is large relative to unity, the criterion is approximately

$$e_1 \leqq R_e E^*/\mu R. \quad (45)$$

$R_e E^*/\mu R$, however, is the limit of conductive operation for the step function. This should be the case since (45) is correct in the limit as S approaches infinity. This merely states that, as the rate of change of the input voltage approaches that of the step function, the criterion for conductive operation must also approach that for the step function.

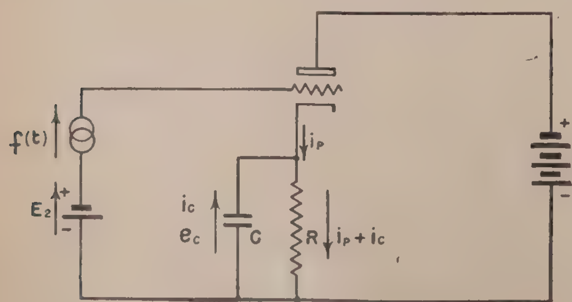


Fig. 6.

If this limit is exceeded, the tube will not conduct over part of the cycle and the output voltage will be exponential over the nonconducting interval. The oscillograms in Fig. 5 show the behavior for a fixed-input sinusoidal voltage as E^* is gradually increased. It is evident that all of the conclusions arrived at for the step function still apply.

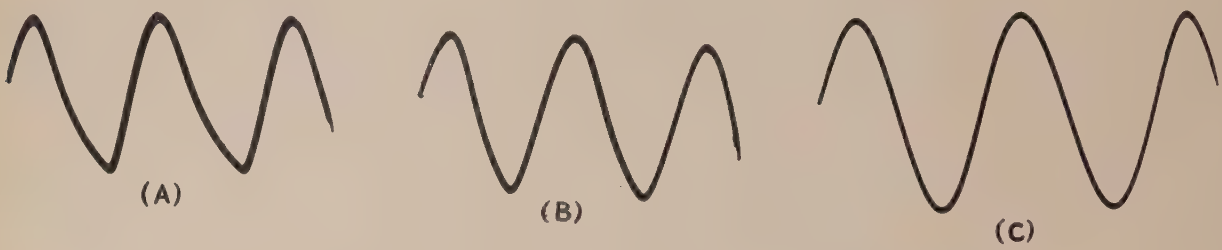


Fig. 5—Results of increasing E^* for sinusoidal input. (E^* increasing to right.)

In the case of the "infinite-impedance" detector, it is necessary that the operation be nonconductive for all sinusoidal input voltages. This may be done by making E^* zero or R infinite. In practice, R is made large and E^* is simply E_b. The detector will not work for vanishingly small inputs, however, or will distort on 100 per cent modulation unless the operation is truly nonconductive. This may be accomplished by making E^* zero which requires that E_2 be equal to $-E_b/\mu$. This is equivalent to saying that the detector should be biased to cutoff. The fact that the ideal triode equation does not apply very close to cutoff modifies this statement somewhat.

Conclusions

The internal impedance of the cathode follower is effective for increasing- as well as decreasing-input voltages as long as the tube conducts. The criterion for conductive operation is that for all times (see Fig. 6)

$$-(\mu+1)e_c + \mu E_2 + \mu f(t) + E_b \geqq 0. \quad (46)$$

In general, as long as the applied voltages to the tube are increasing, this criterion is satisfied. When the applied voltages are decreasing, however, (46) will not be satisfied unless certain precautions are taken and the tube may be rendered nonconducting. When conducting, the cathode follower presents a source impedance equal to $1/(G_m+G_p)$.

The conductive range for applied voltages which are decreasing functions of time may be increased by decreasing the value of R, decreasing the G_m, or increasing the value of E^*. These may be done singly or in any combination depending on circumstances. Where it is not possible to satisfy the conditions for conduction, these measures will improve operation if they are carried as far as possible in any given application.

The choice of a tube for a given cathode-follower application should not be made solely on the basis of G_m, but should be also dictated by the factors analyzed in this paper.

Note on the Fourier Series for Several Pulse Forms*

WILLIAM J. LATTIN†, ASSOCIATE, I.R.E.

Summary—Fourier-series expressions for symmetrical rectangular, triangular, and trapezoidal pulses are derived in a form from which general curves of the magnitudes of the harmonic-amplitude coefficients may be plotted. From these curves it is possible to obtain the values of the amplitude coefficients of the harmonics for any ratio of pulse duration t to cycle period T.

THE FOURIER series for the rectangular pulse, shown in Fig. 1, is as follows:

$$f(t) = E_m\left(\frac{t_1}{T} + \sum_{k=1}^{k=\infty} \frac{2}{\pi k} \sin\ k\frac{t_1}{T}\pi \cos k\omega t\right)$$

$k = 1, 2, 3, 4$, etc.

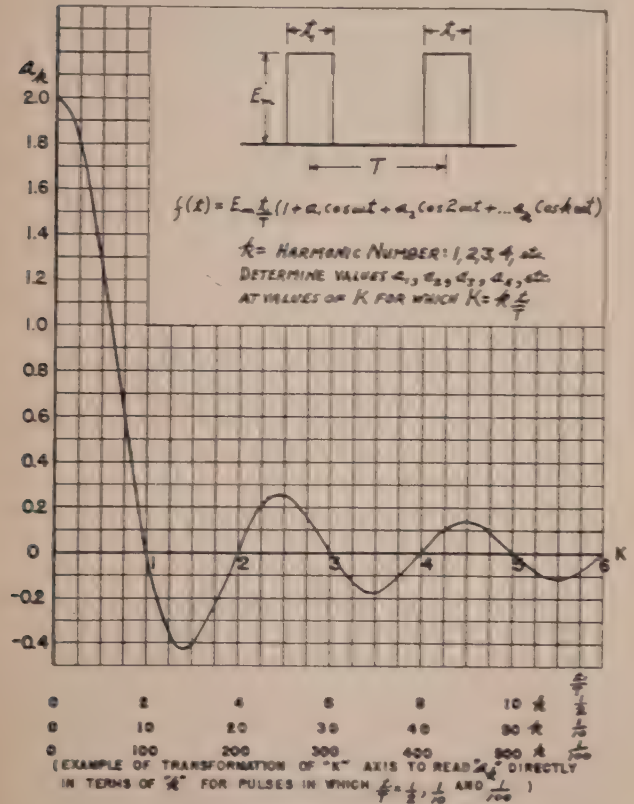


Fig. 1—The rectangular pulse.

The maximum amplitudes of the harmonic components k may be determined if t_1/T is known by evaluating $(2/\pi k) \sin k(t_1/T)\pi$. With the Fourier series in this form, the computation must be made for each harmonic for which one desires to know the amplitude. While a curve of the amplitude coefficients may be plotted versus the harmonic order k for particular values of t_1/T, it obviously will not apply for other values of t_1/T unless we confine the argument to the case where $t_1/T \ll 1$.

By putting the function in a slightly different form, it is possible to arrive at a curve which has no such confinement and enables one to visualize the shape of the plot of amplitude coefficients versus harmonic order for a rectangular pulse having any given value of t_1/T. This is accomplished by substituting $k = K(T/t_1)$ and restricting K to those values for which $k = 1, 2, 3, 4$, etc. Making this substitution, the following expression results:

$$f(t) = E_m\frac{t_1}{T}\left(1 + \sum_{k=1}^{k=\infty} \frac{2}{\pi K} \sin K\pi \cos k\omega t\right),\ K = k\frac{t_1}{T}.$$

The plot of $(2/\pi K) \sin K\pi$ versus K is shown in Fig. 1. From this curve, the harmonic amplitudes may be found for any value of the ratio t_1/T for any harmonic k by reading the values a_k for the point on the K axis where $K = k(t_1/T)$. If the absolute value of the amplitude is required, the figure determined from the curve must be multiplied by $E_m(t_1/T)$. However, in most instances, we are interested in the ratios of harmonic to fundamental amplitudes, and these may be readily determined from the curve. Also, the K axis may be transformed to correspond to the value of t_1/T desired as is shown in Fig. 1 for various values, and the harmonic amplitudes a_k may be read directly from the curve in terms of k. The curve is a general expression, and may be used with any value of t_1/T for which the harmonic amplitudes are required.

As an example, the equation for the so-called square wave $(t_1/T = 1/2)$ may be obtained. For the fundamental frequency, $k = 1$, $K = 1\cdot\frac{1}{2} = 0.5$, and from the curve $a_1 = 1.27$, similarly

$k = 2.0$	$K = 1.0$	$a_2 = 0$
$k = 3.0$	$K = 1.5$	$a_3 = -0.42$
$k = 4.0$	$K = 2.0$	$a_4 = 0$
$k = 5.0$	$K = 2.5$	$a_5 = 0.25$
$k = 6.0$	$K = 3.0$	$a_6 = 0$
$k = 7.0$	$K = 3.5$	$a_7 = -0.18.$

Placing these values in the series, we obtain the square-wave equation

$$f(t) = E_m\cdot(1/2)(1 + 1.27 \cos \omega t - 0.42 \cos 3\omega t + 0.25 \cos 5\omega t - 0.18 \cos 7\omega t + \cdots).$$

The Fourier series for the symmetrical triangular pulse shown in Fig. 2 has been given similar treatment, resulting in the following expression:

$$f(t) = E_m\frac{t_1}{T}\left(\frac{1}{2} + \sum_{k=1}^{k=\infty} \frac{2}{\pi^2 K^2}(1 - \cos K\pi) \cos k\omega t\right),$$

$$K = k(t_1/T).$$

The function $(2/\pi^2K^2)(1 - \cos K\pi)$ versus K is plotted in Fig. 2, and may be used in exactly the same manner for a triangular pulse, as explained above for the rectangular pulse and Fig. 1.

* Decimal classification: 510×537.7. Original manuscript received by the Institute, March 9, 1945; revised manuscript received, June 13, 1945.

† 1304 Locust Street, Owensboro, Kentucky.

The Fourier series for the symmetrical trapezoidal pulse shown in Fig. 3 is

$$f(t)=E_m\frac{t_2}{T}\left(\frac{1+r}{2}+\sum_{k=1}^{k=\infty}\frac{2(\cos rK\pi-\cos K\pi)}{\pi^2K^2(1-r)}\cos k\omega t\right)$$

$$K=k(t_2/T),\qquad r=t_1/t_2.$$

The function $[2(\cos rK\pi-\cos K\pi)/\pi^2K^2(1-r)]$ versus K results in a family of curves for the different values of $r=t_1/t_2$ required. This plot is given in Figs. 3 and 4 for several values of $r=t_1/t_2$. From these curves, at least an estimate of the amplitude coefficients may be obtained for any value of r. Specific values can, of course, be obtained only for the values of r, given on the curves.

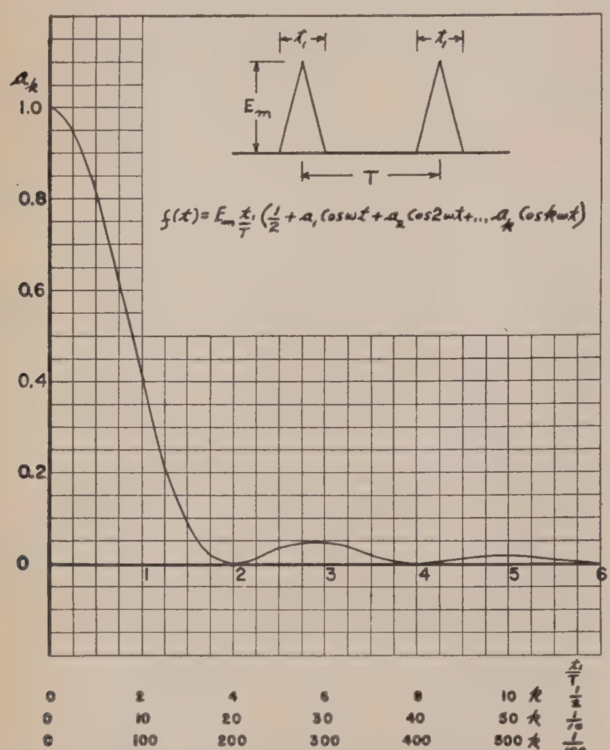


Fig. 2—The symmetrical triangular pulse.

It may be remarked that the triangular pulse is a special case of the above equation; for if $t_1=0$

$$f(t)=E_m\frac{t_2}{T}\left(\frac{1}{2}+\sum_{k=1}^{k=\infty}\frac{2}{\pi^2K^2}(1-\cos K\pi)\cos k\omega t\right).$$

With this information, a complete visual picture may be obtained of the harmonic content of any of the above straight-sided symmetrical pulses. The progression of the curves from the rectangular pulse through the trapezoidal form to the triangular pulse should be noted, since they all belong to the same family of symmetrical pulses, and the general shape of each curve may be more easily remembered.

This treatment may, of course, be applied to other wave forms, if the coefficients in the Fourier series contain the quantity $k\times(t_1/T)$ as a simple product. In wave forms having expressions such as $\sin(k-1)(t_1/T)\pi$, $\sin(k+1)(t_1/T)\pi$, etc., it does not appear feasible.

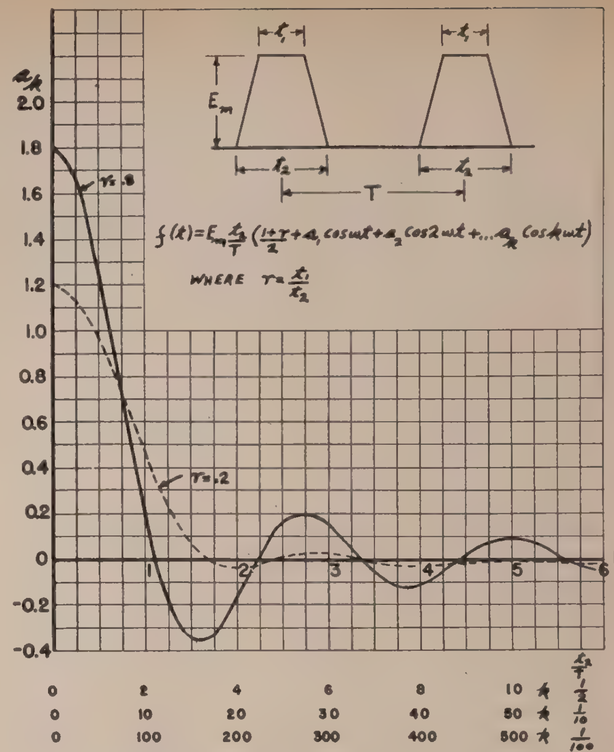


Fig. 3—The symmetrical trapezoidal pulse.

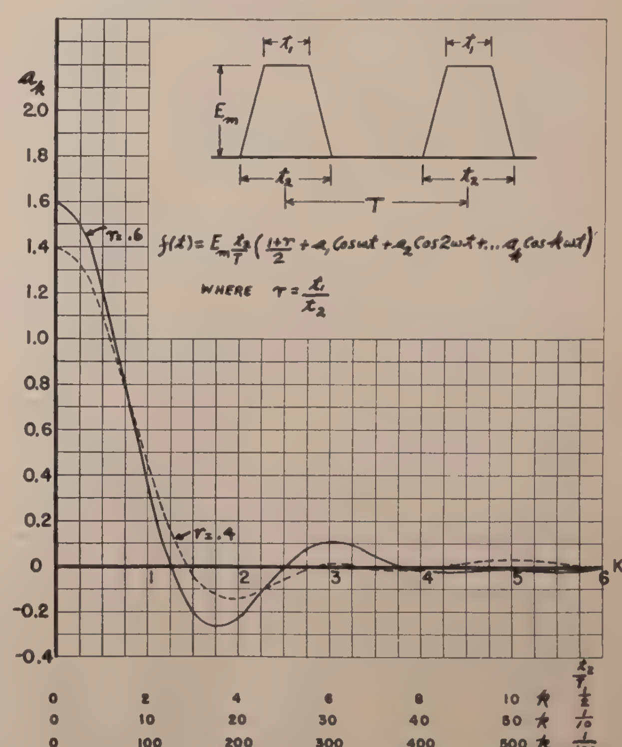


Fig. 4—The symmetrical trapezoidal pulse.

Analysis of Current-Stabilizer Circuits*

W. R. HILL, Jr.†, MEMBER, I.R.E.

Summary—**The performance of any current stabilizer can be predicted in terms of two parameters defined as the stabilization transconductance, g_s, and the output conductance, g_0. Together with the equivalent circuits of Figs. 2 and 3 these two factors permit the calculation of the stabilizer performance in conjunction with any load circuit and direct-current supply. Fundamental stabilizer circuits based on the degenerative and mu-bridge principles are developed and analyzed for the two parameters defined. For simple circuits, the analysis suggests the use of pentodes to obtain best stabilization. Superior performance can be obtained by the mu-balance circuit described. This circuit provides an output current substantially independent of any input-voltage or load-circuit change.**

Introduction

Principal Symbols

Instantaneous Values of Alternating Components

e_c = glow-tube plate-to-cathode voltage
e_g = grid-to-cathode voltage
e_p = plate-to-cathode voltage
e_r = stabilizer output voltage
e_s = stabilizer input voltage
i_p = plate current
i_r = stabilizer output current
i_s = stabilizer input current

Instantaneous Total Values

e_i = stabilizer input voltage
e_0 = stabilizer output voltage
i_i = stabilizer input current
i_0 = stabilizer output current

Effective Values of Alternating Components

E_g = open-circuit rectifier supply voltage
E_r = stabilizer output voltage
E_s = stabilizer input voltage
I_r = stabilizer output current
I_s = stabilizer input current

Average Values

E_c = glow-tube plate-to-cathode voltage
E_i = stabilizer input voltage
E_0 = stabilizer output voltage
I_i = stabilizer input current
I_0 = stabilizer output current

Parameters

g_s = stabilization transconductance, mhos
g_0 = stabilizer output conductance, mhos
R_c = glow-tube dynamic resistance, ohms
r_p = plate resistance, ohms
R_L = load resistance, ohms
Z_g = internal impedance of supply rectifier, ohms
Z_L = load impedance presented to stabilizer, ohms

* Decimal classification: R338. Original manuscript received by the Institute, November 20, 1944; revised manuscript received, January 29, 1945. Reprints of this paper may be obtained without cost by addressing the Engineering Experiment Station, University of Washington, Seattle 5, Washington.
† University of Washington, Seattle 5, Washington.

Z_0 = equivalent impedance presented to load by stabilizer and direct-current supply, ohms
μ = grid-plate amplification factor
μ_{sg} = screen-plate amplification factor
ω = angular velocity, radians per second

General Analysis

Developments of Equivalent Circuits

THE OUTPUT current of a properly designed current-stabilizer circuit is an approximately linear function of the input voltage and the output voltage as expressed by the following relation

$$i_0 = a + be_i + ce_0 \tag{1}$$

where a, b, and c are constants. The following analysis, based upon the use of this equation, is similar to derivation of the equivalent circuit for the vacuum-tube amplifier which gives results that are correct for small-signal amplification and which are useful even when the signal is so large that the vacuum-tube parameters can no longer be considered as constants.

Constant a of (1) is of no interest in predicting the stability of the stabilizer under changes of input voltage or load. The nature of b can be determined by taking the partial derivative of i_0 with respect to e_i. This gives

$$b = \partial i_0/\partial e_i = g_s \tag{2}$$

which will be defined as the stabilization transconductance of the circuit. This factor is a measure of the effectiveness of the circuit in preventing input-voltage changes from affecting the output current.

Taking the partial of i_0 with respect to e_0 gives

$$c = \partial i_0/\partial e_0 = -g_0 \tag{3}$$

which will be defined as the output conductance of the circuit. The negative sign accounts for the fact that an increase of e_0 normally produces a decrease of i_0 which is imagined to be produced by the effect of a positive internal resistance.

Placing the newly defined factors in (1) gives the following expression:

$$i_0 = a + g_se_i - g_0e_0. \tag{4}$$

From this relation, evidently the best stabilizer is one having the smallest possible values of g_s and g_0. In fact, if both factors are zero the output current will be independent of both input- and output-voltage changes.

In stabilizer analysis the changes of voltage and current are of interest. Consequently each of the variables of (4) will be considered to consist of a steady-state value and a number of alternating components. By use of the superposition theorem the analysis can then be carried out for one component alone in a fashion similar to amplifier analysis. Rewriting (4) in terms of a steady-state value and a single alternating component,

$$I_0 + I_{rm} \sin \omega t = a + g_s E_i + g_s E_{sm} \sin (\omega t + \theta_s) - g_0 E_0 - g_0 E_{rm} \sin (\omega t + \theta_r). \quad (5)$$

Under steady-state conditions the alternating components are zero so that $I_0 = a + g_s E_i - g_0 E_0$. Therefore,

$$I_{rm} \sin \omega t = g_s E_{sm} \sin (\omega t + \theta_s) - g_0 E_{rm} \sin (\omega t + \theta_r).$$

Rewriting this in terms of effective vector values,

$$I_r = g_s E_s - g_0 E_r. \quad (6)$$

Fig. 1 shows a schematic diagram of current stabilizer, load impedance, and supply voltage. By Thevenin's theorem the voltage source is represented as a single voltage and internal impedance for each component of

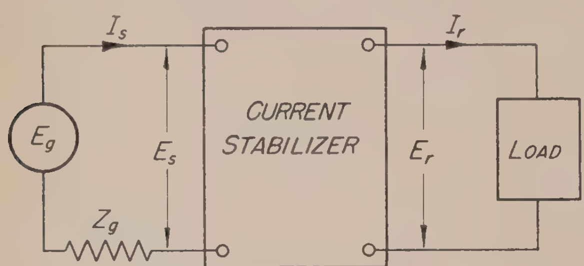


Fig. 1—Schematic diagram of stabilizer and load.

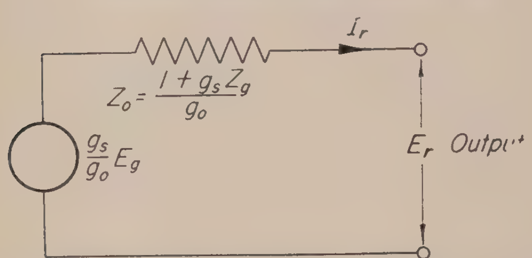


Fig. 2—Equivalent circuit of stabilizer (Thevenin's theorem).

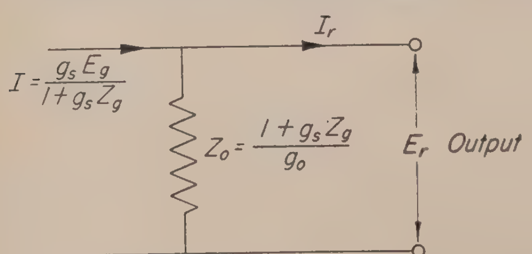


Fig. 3—Equivalent circuit of stabilizer (Norton's theorem).

the input-voltage variations. For slow variations of input voltage Z_g is equal to the slope of the voltage-current characteristic of the rectifier. For high-frequency variations Z_g is nearly equal to the reactance of the filter output capacitor. E_g is the open-circuit value of the rectifier-voltage variations.

Reference to Fig. 1 shows that $E_s = E_g - I_s Z_g$. In many stabilizer circuits the difference between the input and output currents is negligibly small, such that $I_s = I_r$. Although this is not true for some circuits, it can be made exact by considering impedances shunting the stabilizer input terminals as part of Z_g. Consequently $E_s = E_g - I_r Z_g$. Substituting this value for E_s in (6),

$$I_r = g_s E_g - g_s I_r Z_g - g_0 E_r.$$

Solving for E_r,

$$E_r = (g_s E_g / g_0) - I_r((1 + g_s Z_g)/g_0). \quad (7)$$

Equation (7) can be represented by the equivalent circuits shown in Figs. 2 and 3. Fig. 2 shows the equivalent circuit (Thevenin's theorem) to consists of a voltage $g_s E_g / g_0$ in series with an impedance $Z_0 = (1 + g_s Z_g)/g_0$. Since the quantity $g_s Z_g$ is usually small compared to unity, the internal impedance Z_0 is nearly equal to $1/g_0$ or r_0.

When the output conductance g_0 approaches zero the circuit of Fig. 2 becomes indeterminate, making it more convenient to use the equivalent constant-current circuit shown in Fig. 3 (Norton's theorem).

Use of Equivalent Circuit to Predict Stabilizer Performance

A current stabilizer attempts to supply a constant current regardless of any change in input voltage or output voltage. In many cases a change of load resistance causes the change in output voltage. The performance with respect to changes in input voltage can easily be predicted with the help of Fig. 2. In this circuit the current change I_r due to an input-voltage change E_g is

$$I_r = g_s E_g / g_0 (Z_0 + Z_L) = [g_s E_g / g_0 (Z_L + (1 + g_s Z_g)/g_0)]. \quad (8)$$

In using the relation given by (8) it must be remembered that Z_L and Z_g are functions of the frequency so that the performance of the circuit may depend upon the frequency of the input-voltage variations. For instance, when analyzing the system for slow input-voltage changes, Z_g is practically equal to the direct-current internal resistance of the supply rectifier (the slope of the voltage-current characteristic curve) and Z_L is the direct-current resistance of the load. For analysis of the effect of the rectifier-ripple voltage on the current output, however, Z_g is practically equal to the reactance of the filter output capacitor and Z_L must be taken as the impedance of the load to the ripple frequency.

As an example of the use of (8) assume that, for the stabilizer under investigation, g_0 is 20 micromhos and g_s is 10 micromhos. The stabilizer is to be used with a rectifier having a direct-current internal resistance of 1000 ohms, and a load having a direct-current resistance of 1000 ohms. Computing the internal impedance of the system,

$$Z_0 = (1 + (10^{-5})(1000))/(2)(10^{-5}) = 50{,}500 \text{ ohms}.$$

If a line-voltage variation of 10 per cent is expected and the rectifier output voltage without load is 300 volts, the open-circuit change in rectifier output voltage will be approximately 30 volts. Inserting this information into (8) the resulting change in output current can be computed

$$I_r = 10(30)/(20)(50{,}500 + 1000) = 0.00029 \text{ ampere} = 0.29 \text{ milliampere}.$$

The performance of the circuit with respect to the rectifier-output ripple would be substantially the same because in this particular case the values of Z_g and Z_L affect the problem by only a few per cent.

The effect of output-voltage changes on the load current can be determined from an inspection of Fig. 2.

$$I_r = -E_r/Z_0 = -E_r/((1+g_sZ_0)/g_0). \quad (9)$$

The negative sign takes account of the fact that an increase of E_r causes a decrease in output current. For the stabilizer just considered a 30-volt change in output voltage would cause a change of 30/(50,500) or 0.59 milliampere in output current.

A calculation of the effect of a change in load resistance on the output current requires a use of the fact that the stabilizer output consists of a steady-state value of current in addition to the current I_r shown in Fig. 2. The analysis need be carried out only for slow changes in load resistance since high-speed alternating changes in resistance are seldom encountered. Under the initial steady-state conditions the output voltage E_0 produced by current I_0 flowing through R_L is

$$E_0 = I_0R_L. \quad (10)$$

An increase ΔR_L in the load produces changes E_r and I_r in the output voltage and current. Under these conditions $E_0 + E_r = (I_0 + I_r)(R_L + \Delta R_L)$.
Expanding and subtracting (10),

$$E_r = I_0\Delta R_L + I_rR_L + I_r\Delta R_L. \quad (11)$$

This gives the value of output-voltage change E_r produced by the change in load resistance. Substituting the value of E_r from (11) into (9) and solving for I_r,

$$I_r = -I_0\Delta R_L/(Z_0 + R_L + \Delta R_L). \quad (12)$$

For the stabilizer used as an example, the change in a load current of 50 milliamperes produced by changing the load resistance from 1000 ohms to 1500 ohms would be

$$I_r = -50(500)/(50{,}500 + 1500) = 0.48 \text{ milliampere.}$$

Effect of Frequency on g_s and g_0

Strictly speaking, factors g_s and g_0 should be defined as admittances y_s and y_0 to take account of the fact that there may be a phase angle between E_r and I_r or E_s and I_r. For stabilizer circuits involving only vacuum tubes and resistors, however, the two factors will be essentially pure conductances from zero frequency up through and beyond the audio-frequency band depending on the care taken to reduce circuit capacitance. For circuits employing glow tubes, the frequency range is greatly reduced and the factors may have appreciable reactive components at frequencies below 1000 cycles per second. This is caused by the relatively slow transit time and recombination rate of the positive ions in the discharge. The effect of this is to make the glow-tube impedance appear to have an inductive component; if this is measured its effect on the circuit performance can be estimated or computed by means of the expressions set forth in the following analysis.

Analysis of Specific Circuits

The following discussion is concerned with the analysis of typical stabilizer circuits for the factors g_s and g_0. To aid in gaining an understanding of the operation of the various stabilizer circuits, simple basic types will be analyzed first.

Basic Degenerative Stabilizer

The basic degenerative-type stabilizer is shown in Fig. 4. The operation of this circuit depends on the fact that any change in load current changes the grid bias of V_1 in such a direction as to oppose that load-current change.

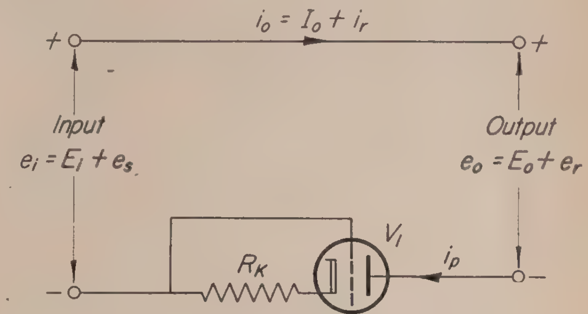


Fig. 4—Basic degenerative stabilizer.

Fig. 4 will now be analyzed for the two factors g_s and g_0. Writing (2) of the general analysis in terms of alternating components,

$$g_s = \partial i_0/\partial e_i = i_r/e_s\,|_{e_r=0}. \quad (13)$$

A change in load current i_r will produce a change in grid voltage of $e_g = -i_rR_k$, while a change of e_s at the input with e_0 held constant will cause a change in the plate voltage of V_1 equal to $e_p = e_s - i_rR_k$. Substituting these values for e_g and e_p in the basic vacuum-tube relation

$$i_p = (\mu e_g + e_p)/r_p$$

we obtain $i_r = i_p = (-\mu i_rR_k + e_s - i_rR_k)/r_p$.
Solving this relation for the ratio i_r/e_s

$$g_s = i_r/e_s = 1/(r_p + (\mu+1)R_k). \quad (14)$$

The factor g_0 as defined in (3) can be written in terms of alternating components as

$$g_0 = -i_r/e_r\,|_{e_s=0}. \quad (15)$$

Examination of Fig. 4, however, shows that the sources of input and output voltage and the stabilizing circuit comprise a simple series circuit. Consequently, an increase in input voltage will have exactly the same effect on the current as a decrease in output voltage. As a result

$$g_0 = -i_r/e_r\,|_{e_s=0} = -i_r/-e_s\,|_{e_r=0} = g_s. \quad (16)$$

Consequently, for a circuit of this type in which no shunt impedances appear across the input or output, the factors g_s and g_0 are equal in magnitude.

Good stabilization requires a small g_s and g_0 so that V_1 should possess a high μ and r_p. This suggests the choice of a pentode, although this entails additional circuit complications. The transconductance of the tube is of secondary importance, although from the standpoint of minimum voltage drop in the stabilizer a high transconductance is desirable. Resistance R_k should be large, but in this simple circuit the value will be dictated largely by the choice of tube and operating voltages.

Basic Mu-Bridge Stabilizer

An improvement in the stabilization transconductance g_s can be obtained with a circuit based on the familiar bridge circuit for measuring amplification factor. However, no improvement in factor g_0 is obtained. This circuit is shown in Fig. 5. Operation of this circuit takes advantage of the fact that two opposing voltage changes applied to the grid and plate, one mu times the other, results in no plate-current change. In Fig. 5,

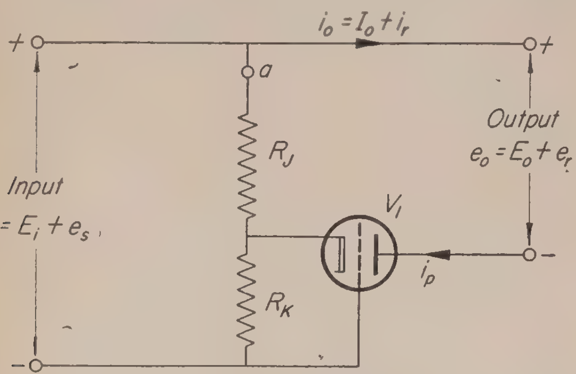


Fig. 5—Basic mu-bridge circuit.

resistor R_j is made equal to μR_k and any change e_s in the input voltage is divided and applied to both plate and grid of V_1 with no resultant change in plate current. Owing to R_k in the cathode circuit the stabilizer is a combination of mu-bridge and degenerative types, and any change in the load circuit is corrected for as effectively as with the circuit of Fig. 4.

To obtain g_s for Fig. 5 the circuit must be analyzed for the ratio i_r/e_s with e_r held zero as defined in (13). The grid-cathode voltage of V_1 is a function of both e_s and i_r.

$$e_g = -i_r R_k' - U e_s$$

where

$$R_k' = R_j R_k/(R_j + R_k),$$

and

$$U = R_k/(R_j + R_k).$$

This is obtained by applying the superposition theorem and determining the separate effects of i_r and e_s before combining them. The plate-cathode voltage of V_1 can be written in a similar manner as $e_p = e_s - i_r R_k' - U e_s$. The values e_p, e_g, and i_p of V_1 are related by the expression

$$i_p = i_r = (\mu e_g + e_p)/r_p.$$

Substituting for e_g and e_p their equivalent expressions and solving for the ratio i_r/e_s,

$$g_s = i_r/e_s = (1 - (\mu + 1)U)/(r_p + (\mu + 1)R_k'). \quad (17)$$

To obtain complete independence from supply voltage variations, g_s must be zero which means that the numerator of (17) must equal zero. Performing this operation, solving for U, and replacing U in terms of and R_k, $U = 1/(\mu + 1) = R_k/(R_k + R_j)$.

Solving for R_j,

$$R_j = \mu R_k \quad (18)$$

as had been anticipated. Since the parameter μ is not strictly a constant, this relation cannot be satisfied for all values of load current and the numerator of (17) will not remain exactly zero. Consequently, the inclusion of the degenerative effect of R_k in the circuit is of value. This is indicated in (17) by the term $(\mu+1)R_k'$ which increases the denominator.

The behavior of the circuit with respect to output-voltage variations is no better than that of Fig. 4. Inspection of the circuit shows that the only difference between an input- and an output-voltage change is that an output change has no effect on the grid voltage of V_1 other than through a change in load current. This amounts to repeating the derivation of g_s with $-e_r$ replacing e_s and with factor U equal to zero. Making this change the expression for g_0 is obtained.

$$g_0 = 1/(r_p + (\mu + 1)R_k').$$

Replacing R_k' by the expression $R_j R_k/(R_j + R_k)$ and remembering that $R_j = \mu R_k$ if the circuit is properly designed, the expression for g_0 becomes

$$g_0 = 1/(r_p + \mu R_k). \quad (19)$$

This represents a slight sacrifice in the value of g_0 compared to (14) to obtain a great improvement in g_s.

Although the circuit of Fig. 5 is a convenient one for illustrating the basic mu-bridge circuit it is not a practical one because the direct current flowing through R_j would produce a voltage drop in R_k that would bias V_1 approximately to cutoff. A practical solution is to insert a constant-voltage glow tube at point a in the circuit. This reduces the current through R_j and R_k and permits V_1 to pass current. As far as changes in input voltage are concerned, however, the glow tube has a negligible effect on the operation of the circuit. Other circuit improvements to permit easy adjustment of the output current will suggest themselves after a study of the circuit.

Pentode Stabilizer (Mu-Bridge Circuit)

The expressions representing g_0 and g_s for the two basic circuits analyzed suggest the use of pentodes in place of triodes in order to obtain high values of r_p and u, at the same time maintaining a high transconductance and low voltage drop in the stabilizing circuit. Comparison of pentodes and triodes of equivalent current rating indicates an advantage of the order of 100 to 1 for the pentode. Utilization of a pentode, however, complicates the circuit because of the necessity of providing a constant screen potential. A practical solution to the problem of inserting a pentode into the basic circuit of Fig. 5 is shown by Fig. 6.

Substantially constant voltage for the screen of V_1 is obtained by glow tube V_2 (several in series if sufficient voltage is required). A portion of this same constant voltage drop is applied to the control grid through resistors R_2 and R_3, thus making adjustment of grid bias and consequently the output current independent of the value of R_k. This allows the use of large values of R_k in which the voltage drop may be even greater than the

cutoff bias for V_1. Resistor R_j provides compensation for input-voltage changes exactly as in Fig. 5.

A complete solution for this circuit requires a knowledge of the control-grid-plate, screen-plate, control-grid-screen, and plate-screen mu factors, the plate and screen dynamic resistances, and the dynamic resistance of V_2. The solution thus obtained is undesirably complex and, furthermore, no information is published on the majority of the required circuit parameters. This makes it desirable to adopt an approximate solution. Unfortunately, neglecting any one of the factors causes considerable error, but a study of the situation indicates that a useful simplification consists of neglecting the effect of the screen current. It is not, however, permissible to neglect the effect of the screen on the plate current or the variation in voltage drop across V_2. Although this approximate solution may be in error as much as 50 per cent in predicting the values of g_s and g_0, it is at least of value in designing a circuit to obtain the most effective stabilization.

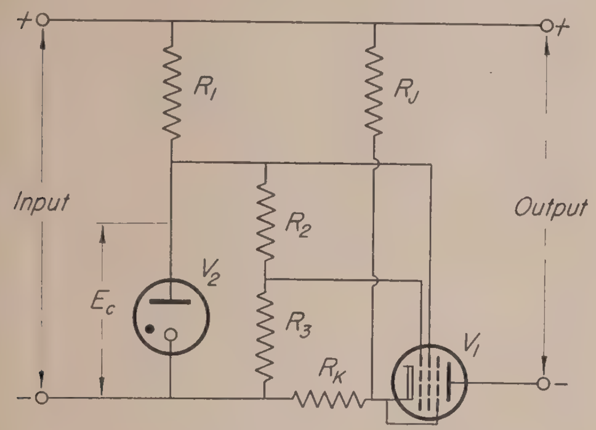


Fig. 6—Pentode stabilizer (mu-bridge circuit).

The solution based on these approximations is[1]

$$g_s \cong \frac{1+\left(\dfrac{R_c}{R_1+R_c}\right)\left[\dfrac{\mu R_3}{R_2+R_3}+\mu_{sg}\right]-\left(\dfrac{R_k}{R_k+R_j}\right)(\mu+\mu_{sg})}{r_p+R_k(\mu+\mu_{sg})}. \quad (20)$$

To make g_s zero it is necessary to set the numerator of (20) equal to zero. Performing this operation and solving for the ratio R_j/R_k,

$$\frac{R_j}{R_k} \cong \frac{\mu+\mu_{sg}}{1+(R_c/(R_1+R_c))\left[\mu R_3/(R_2+R_3)+\mu_{sg}\right]}-1. \quad (21)$$

This determines R_j once R_k has been chosen. The value of R_1 is determined by the operating voltage and current for V_2, R_c is the dynamic resistance of V_2, and the ratio between R_2 and R_3 is determined by the bias needed on V_1.

[1] For frequencies where the impedance of the glow tube has an appreciable reactive component, this complex impedance should be substituted for the R_c shown. The use of impedance requires that the analysis be carried out in terms of effective instead of instantaneous alternating values. The resulting expression is the same in either case.

The corresponding solution for g_0 is

$$g_0 \cong 1/(r_p+R_k(\mu+\mu_{sg})). \quad (22)$$

The complete circuit analysis indicates that the effect of screen current is to make g_s smaller than given by (20) and g_0 larger than given by (22).

It should be noted that the only difference between the mu-bridge pentode circuit and a straight degenerative circuit is the inclusion of resistor R_j. Consequently g_s and g_0 for the degenerative pentode circuit can be obtained by substituting $R_j=\infty$ in (20) and (22).

In order to illustrate the use of Fig. 6 and (20), (21), and (22), a stabilizer design will now be carried out and compared to the corresponding experimental circuit. The circuit will be designed for the following assumed operating conditions: $E_0=150$ volts; $I_0=30$ milliamperes; $E_i=400$ volts. A tube suitable for this output is a 6F6 pentode for which $\mu=180$, $\mu_{sg}=20$, and $r_p=70{,}000$ ohms. For a screen voltage of 200 volts, E_c must be 200 volts plus the drop in R_k. A *VR* 150/30 in series with a *VR* 105/30 makes a satisfactory combination. For the particular glow tubes used, the total value of R_c was 600 ohms. A convenient value for R_k is 1000 ohms. The load current plus an estimated screen current of 5 milliamperes and current through R_j of about 10 milliamperes produce a total drop in R_k of 45 volts. This makes the screen voltage of V_1 about 200 volts as planned. From the tube characteristics the grid bias required on V_1 for a 30-milliampere plate current is about -12 volts. Thus the values of R_3 and R_2 must be such that the drop across R_3 is $45-12=33$ volts. The remainder of the 250-volt E_c must be across R_2; hence $R_2/R_3=(250-33)/33=6.6$. R_2 and R_3 should be as large as possible except that R_3 must not exceed the maximum grid-circuit resistance allowed for V_1. It is convenient to make R_3 adjustable to permit adjustment of the output current. Resistor R_1 must provide a drop of 150 volts while carrying a glow-tube operating current of 15 milliamperes and a screen current of 5 milliamperes. Its value is then 7500 ohms.

The value of R_j can now be computed from (21). Inserting the various values already determined gives a value of 45,000 ohms for R_j. A check should now be made to correct the estimate of the current through R_j used in establishing the voltage drop across R_k. This check shows the R_j current to be 8 milliamperes instead of the assumed 10 milliamperes which is close enough to make recomputation unnecessary. Extreme care in carrying out the computations is not justified by the approximations made in deriving (21).

The performance of the circuit designed above was tested in the laboratory with the results shown by the curves of Figs. 7 and 8. Since the slope of the curve in Fig. 7 is the factor g_s, resistor R_j was adjusted until the curve was flat at the assumed input voltage of 400 volts. The actual value of R_j required to establish zero g_s at this point was 39,000 ohms as compared with the

computed value of 45,000. The difference is an indication of the inaccuracy to be expected from the simplifications made in obtaining (21).

The negative slope of the curve of Fig. 8 is a measure of the output conductance g_0 of the circuit. This slope

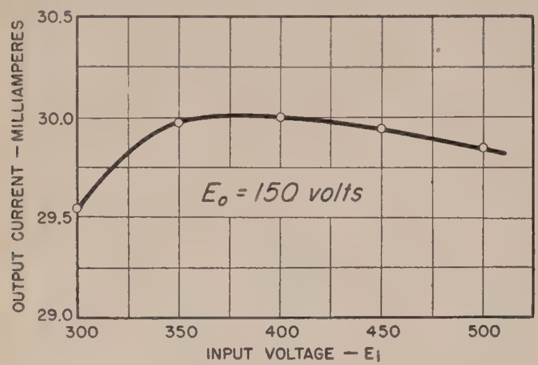


Fig. 7—Performance of pentode mu-bridge circuit.

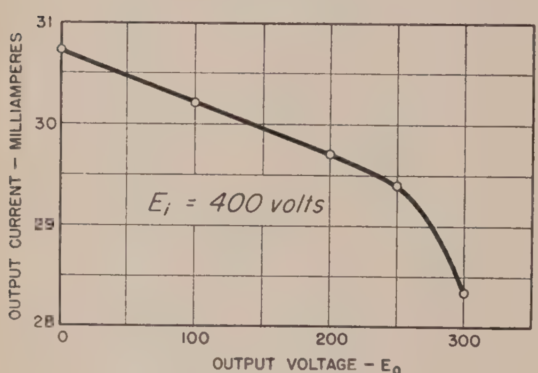


Fig. 8—Performance of pentode mu-bridge circuit.

is equal to 5.3 micromhos at the assumed output voltage of 150 volts. Computing g_0 from (22) a value of 3.7 micromhos is obtained.

The behavior of the stabilizer as a degenerative type with R_j removed is shown in Fig. 9. Comparison with

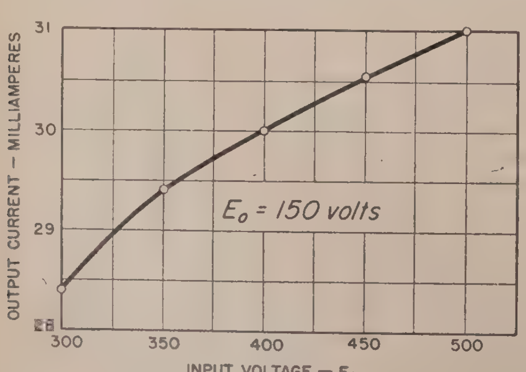


Fig. 9—Performance of degenerative pentode circuit.

Fig. 7 immediately shows the advantage of the mu-bridge circuit. Measurement of the slope at the operating point yields a g_o of 11 micromhos. The value computed from (20) with R_j equal to infinity and the ratio R_2/R_3 adjusted to re-establish the correct bias is 13.5 micromhos. The output conductance g_0 is not appreciably affected by removing R_j.

Mu-Balance Stabilizer

Substantially the same order of stabilization as that obtained with pentodes can be obtained by using a triode control tube in conjunction with an amplifier for amplifying the voltage changes before applying them to the control-tube grid. Use of triodes, however, has the advantage that the characteristic obtained is more nearly straight than when pentodes are employed. As a result the mu-bridge principle is more effective because the balance is maintained over a wider range of voltage changes. For example, in Fig. 7 the curve can be made flat at any point in the operating range desired but, owing to the rapid change of μ and r_p with applied voltage, the over-all regulation for large input-voltage changes is not particularly good.

The addition of an amplifier affords another important advantage; it provides a convenient phase inversion making it possible to apply the mu-balance principle in the way suggested by Fig. 10. With this circuit any

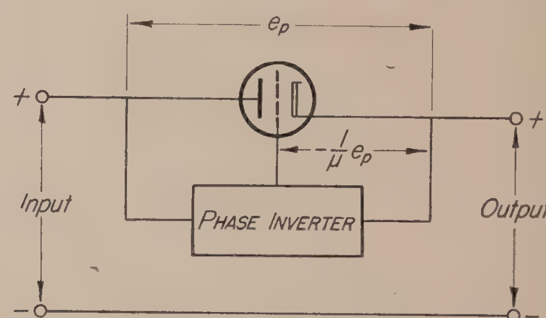


Fig. 10—Basic mu-balance circuit.

change in either input or output voltage which produces a voltage change e_p across the control tube will also provide an inverted voltage change at the grid to maintain the plate-current constant. By employing triodes which have the characteristic of nearly constant amplification factor over most of the operating region, very good stabilization of output current can be obtained. Beside providing better stabilization, circuits based on this principle have the advantage of correcting equally well for both input- or output-voltage changes and they may be placed in either the positive or negative side of the circuit.

In practice it is desirable to supplement the mu-balance circuit with degeneration in order to reduce the effects of any unbalance that may occur over portions of the operating range. A practical stabilization circuit embodying these two principles is shown by Fig. 11. In this circuit, resistor R_k serves to provide a degenerative voltage that is amplified by V_2 and applied to the grid of V_1. The voltage drop in R_k also serves as a means of obtaining a negative grid bias for V_1. Resistors R_2 and R_3 serve to select a fraction of the voltage drop across V_1 for amplification and phase inversion by V_2 as discussed

in connection with Fig. 10. Glow tube V_3 serves as a reference voltage against which the drop in R_k is compared by V_2. Glow tube V_4 (one or more of the one-quarter-watt type) permits applying the plate-voltage changes of V_2 to the grid of V_1 although the two elements must operate at a direct-current difference of potential.

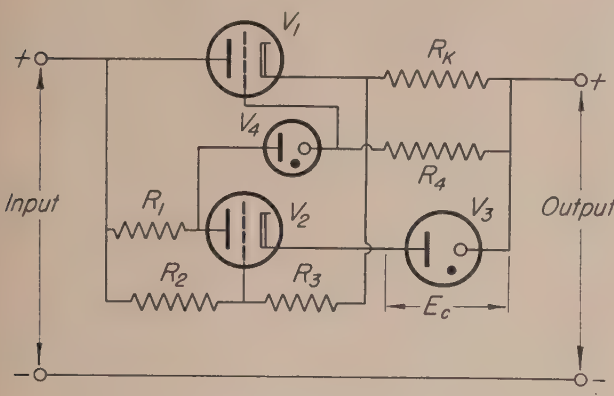


Fig. 11—Degenerative mu-balance circuit.

The performance obtained with the circuit of Fig. 11 is shown by the curve of Fig. 12. A single curve is sufficient to describe the circuit characteristic because the

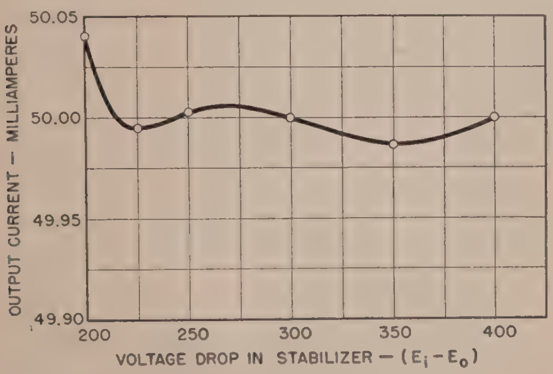


Fig. 12—Performance of degenerative mu-balance circuit.

stabilizer takes the form of a two-terminal network in series with the output. Consequently the current flow depends entirely on the difference between input and output voltages. From a comparison of the curves it is apparent that this circuit is about thirty times more effective than the pentode mu-bridge circuit; furthermore, it is equally effective for both input and output variations, whereas the mu-bridge circuit is not.

The design of the circuit of Fig. 11 is not difficult. Resistors R_1, R_4, and R_k are dictated by the choice of tubes and operating conditions. There remains the choice of R_2 and R_3 which determines the circuit balance and controls the performance of the stabilizer. In practice it is convenient to use a potentiometer for the R_2-R_3 combination and to adjust the circuit for best performance experimentally. This can be done by applying to the input an alternating voltage in addition to the direct voltage and observing the voltage across a load resistance with an oscilloscope. The potentiometer is then adjusted until the oscilloscope figure is of minimum height.

An analysis of the circuit leads to an equation involving the circuit elements that is difficult to solve explicitly for the ratio of R_2 to R_3. By making certain simplifying assumptions, however, it is possible to establish an approximate value for this ratio that is of use in selecting suitable values for R_2 and R_3. The approximate expression is as follows:

$$\frac{R_2}{R_3} \cong \frac{\mu_1\mu_2 R_1}{(1+\mu_1)[\mu_2(R_k+R_c)+r_{p2}]-R_1} - 1. \quad (23)$$

The actual values of R_2 and R_3 should be as high as possible; ordinarily R_2 can be made larger than 1 megohm.

The design procedure for the circuit of Fig. 11 is as follows: The voltage drops across V_3 and R_k differ only by the bias on V_2 and the small drop in R_3; therefore, if V_2 is a high-mu triode operating at a small negative bias, the voltage drop I_0R_k is nearly equal to E_0. R_4 is selected to be as high as possible yet to draw sufficient current through V_4. By employing one-quarter-watt neon lamps for V_4 this operating current can be made less than 100 microamperes. R_1 is chosen as a proper load resistance for V_2 consistent with proper operating plate voltage for V_2. The ratio R_2/R_3 is then computed and values of R_2 and R_3 are chosen.

A sample design for a circuit capable of an output current of 50 milliamperes will serve to illustrate this procedure. Suitable tubes are a triode-connected 6L6 for V_1 and a type-6SF5 high-mu triode for V_2. A 2-watt neon lamp provides a value of E_c of about 60 volts which is a compromise between using a high E_c and correspondingly high R_k with improved degeneration, or a low E_c to reduce the over-all circuit drop. The value of R_k will then be approximately E_c/I_0 or 1000 ohms. The operating voltage across V_1 must next be selected. A value of 200 volts is sufficient for V_1 to pass 50 milliamperes with a reasonably negative grid voltage (-12 volts). The over-all drop in the stabilizer will then be 250 volts under the assumed operating conditions.

R_4 is selected by observing that the drop across it is about 38 volts and that it must carry a current of about 50 to 100 microamperes to operate V_4. A suitable value for R_4 is 500,000 ohms. The plate of V_2 should operate at about 100 volts as a compromise between maximum plate voltage and reasonable allowable drop in R_1. This requires the drop across V_4 to be about 112 volts. Two one-quarter-watt neon bulbs in series provide a drop of 110 volts which fits the circuit requirements nicely. The glow tubes used for V_3 and V_4 must be of the type without series resistance in the base. Reference to the characteristic curves for V_2 indicates that with a plate voltage of 100 volts a reasonable value of plate current is 0.3 milliampere. This will be obtained with a negative grid voltage of one volt, a value well below the region

where appreciable grid current flows. Resistor R_1 must then carry 0.3 milliampere plus the current in R_4 or a total of 0.376 milliampere with a voltage drop of 100 volts; a resistance of 250,000 ohms is required. The current flow through V_3 is rather small but a 2-watt neon lamp is a satisfactory choice. For this tube a value of R_e of 400 ohms was observed.

The ratio R_2/R_3 can now be computed. Inserting the circuit constants into (23) a value of 108 is obtained. For an R_2 of 1 megohm, R_3 should be about 9000 ohms. A convenient arrangement employs a 15,000 potentiometer for R_3.

The circuit described above was tested experimentally with the results shown in Fig. 12. The actual value of the ratio R_2/R_3 required to obtain Fig. 12 was 94 instead of the computed approximation of 108. Other circuit constants were identical to those designed with the exception that adjustment of I_0 to exactly 50 milliamperes required a value of 940 ohms for R_k instead of the estimated 1000 ohms.

Dynamics of Electron Beams*

Applications of Hamiltonian Dynamics to Electronic Problems

D. GABOR†

Survey of the Problem

IT IS THE purpose of this paper to present certain advanced theories of dynamics in a form in which they may be useful to the electronic research worker. The foundations of these theories were mostly laid by Sir William Rowan Hamilton, over a hundred years ago.

Hamilton has left three complete formulations of dynamics, equivalent in meaning, as different as possible in form. These are Hamilton's principle, the canonical equations, and the Hamilton-Jacobi equation. Only the first of these appears to be well known among electronic research workers, as this principle is usually made the starting point of treatises on electron optics. This paper deals, therefore, mainly with the other two, which deserve to be better known and more widely used.

The dynamical problems which may present themselves in electronic devices can be conveniently graded in six stages, each with three subdivisions. There are thus, in all, 18 stages of more or less continuously increasing difficulty:

(A) *The motion of a single electron*, in

1. electrostatic fields } (steady or quasi-
2. static electromagnetic fields } steady motion).
3. variable electromagnetic fields ("transit-time" effects).

The fields can be considered as quasi-static so long as they do not change appreciably during the passage of an electron. We adopt the subdivision 1 to 3 also in the following stages, but do not write it down explicitly.

(B) *"Regular" electron flow*, without random motion, with negligible current and space charge.

(C) *Electron flow with random motion*, but negligible current and space charge.

(D) *Electron flow with space charge*, but replacing the electrons by a continuous charge and current distribution.

(E) *Fluctuations of the electron flow*, due to the particulate nature of the charge carriers (noise).

(F) *Interaction of individual electrons* with one another (degeneration of the regular flow into random motion).

Only the stages A, B, and C will be discussed in this paper, each with 3 subdivisions; that is, nine cases in all. If these are attacked with elementary or *ad hoc* methods, each of them presents a separate problem. It is the fundamental advantage of the advanced Hamiltonian methods that, by applying them, *each problem can be reduced to a simpler type;* i.e., it can be moved one or sometimes several stages down in the hierarchy of difficulties. For instance, we shall be able to treat motion in electromagnetic fields essentially by the same method as in purely electrostatic fields, and transit-time effects by the same methods as problems of steady motion. Moreover, we shall have little to do with the motion of a single electron, as Hamilton's methods enable us to treat the motion of a whole "regular" beam, emitted by a cathode, in a unified fashion. Finally, the problem of random motion can be reduced under very general assumptions to the problem of regular flow.

The Motion of a Single Electron—Hamilton's Canonical Equations

If an electron with charge $-e$ moves in an electromagnetic field with electric-field intensity $\mathbf{E}$ and magnetic intensity $\mathbf{H}$, its law of motion due to Lorentz is

$$\frac{d\mathbf{p}_m}{dt} = -e\left[\mathbf{E} + \frac{1}{c}(\mathbf{v}\times\mathbf{H})\right]. \tag{1}$$

This is the Newtonian form of the law of motion. At the

* Decimal classification: R138. Original manuscript received by the Institute, August 13, 1945. This paper is an expanded version of a lecture before a joint meeting of the Northwest Section of the Institution of Electronics and the Manchester District Branch of the Institute of Physics, April 14, 1944. The lecture was printed in *The Science Forum* (*Electronics Section*), September, 1944.

† The British Thomson-Houston Company, Ltd., Rugby, England.

left side we have the rate of change of the mechanical momentum $\mathbf{p}_m$, which, for velocities small against the velocity of light, is $\mathbf{p}_m = m\mathbf{v}$, where $\mathbf{v}$ is the vector of velocity and m is the mass of the electron. The expression at the right side is the force. The units are Gaussian.[1]

This is a rather complicated and unhandy expression, though it is the one most often used. Hamilton found in 1834 a different formulation of the law of motion. The fundamental idea is to consider the momentum $\mathbf{p}$ not as a derivative of the motion in space, but as a vector in a special "momentum space," with co-ordinates p_x, p_y, p_z. These three monentum co-ordinates together with the "configuration" co-ordinates x, y, z, entirely describe the momentary dynamical state of the particle. The reason for this doubling of data will become fully manifest somewhat later, in the discussion of random motion. With these six variables the law of motion can be written down in the "canonical" form[2]

$$\frac{dx}{dt} = \frac{\partial \mathcal{H}}{\partial p_x} \qquad \frac{dp_x}{dt} = -\frac{\partial \mathcal{H}}{\partial x}. \tag{2}$$

Corresponding equations obtain for y and z. $\mathcal{H}$ is the "Hamiltonian," defined as the total energy of the particle, *expressed by the position* (x, y, z) *and the momentum*[3] (p_x, p_y, p_z). Hamilton himself did not think of applying these equations to forces of the queer type which act at right angles to the velocity. It was only about 70 years later when it was discovered that the "canonical equations" can be used also for the description of electron motion in electromagnetic fields, if the Hamiltonian is assumed as follows:[4,5]

$$\mathcal{H} = \frac{1}{2m}\left[\left(p_x + \frac{e}{c}A_x\right)^2 + \left(p_y + \frac{e}{c}A_y\right)^2 + \left(p_z + \frac{e}{c}A_z\right)^2\right] - e\phi. \tag{3}$$

In this equation, ϕ is the scalar or "electrostatic" potential, while A_x, A_y, A_z, are the components of the vector potential A. The second term, $-e\phi$ is the potential energy of the electron, hence we must expect that the first term is the kinetic energy. This can be verified by substituting $\mathcal{H}$ into the first set of the canonical equations (2). These give, in vectorial form

$$\mathbf{v} = \frac{1}{m}\left(\mathbf{p} + \frac{e}{c}\mathbf{A}\right). \tag{4}$$

On substituting $\mathcal{H}$ into the second set of (2), these turn out to be equivalent to the law of motion (1).

The essence of this is that the magnetic field and its complicated action on the electron can be taken into account by the simple expedient of *giving a new definition for the momentum* $\mathbf{p}$, which may now be called "total" momentum. This definition is, according to (4)

$$\mathbf{p} = m\mathbf{v} - \frac{e}{c}\mathbf{A} = \mathbf{p}_m - \frac{e}{c}\mathbf{A}.$$

The total momentum of an electron in a magnetic field is the sum of the mechanical momentum $\mathbf{p}_m$ and of the vector potential A, multiplied by a universal constant $-(e/c)$.

It may be noted that the vector potential is not entirely defined. In the same way as a constant can always be added to ϕ, the gradient of an arbitrary scalar funcrion can always be added to $\mathbf{A}$, without affecting the electromagnetic field.[6]

The first obvious advantage of the Hamiltonian form of the equations of motion is, thus, that electron motion in electromagnetic fields can be treated formally (and as we shall see later, also practically), by the same methods as in electrostatic fields. A second advantage is that these equations lead straight to the two fundamental *invariants* of dynamics, the invariants of Lagrange and of Liouville.

Lagrange's Invariant

Quantities which remain constant during the motion of an electron are called *integrals of the motion*. But the well-known integrals of dynamics, the integrals of momentum and of energy, are of rather limited use in electronic devices. The integrals of momentum can be usefully applied only in special cases, distinguished by a certain symmetry, and the integral of energy applies only in static or quasi-static cases, but not in fields which vary appreciably during the transit time of an electron. On the other hand, the two fundamental invariants are of general and unrestricted validity in the electronic problems which have been listed under (A)–(E).

Let us consider an electron beam; for instance, the electrons emitted by a cathode, which can be first visualized as a hail of particles. We now replace this "hail" by a continuous distribution of trajectories, densities, velocities, etc. This may be called the "hydrodynamical picture" of the beam. As in hydrodynamics, we forget the elementary particles which compose the fluid. We are allowed to do this, as in all problems (except those of the types (E) and (F)), the constants of the electrons will always figure only in the combination (e/m), hence it does not matter whether both e and m have definite values, or whether both go to the limit zero, so long as their ratio remains the same. With this understanding

[1] In order to convert the formulas in this paper into the m.k.s. system as used, e.g., by J. A. Stratton, "Electromagnetic Theory," McGraw-Hill, Book Company, New York, N. Y., 1941, change $\mathbf{H}$ into $\mu_0\mathbf{H} = B$ and cancel all factors $1/c$ associated with $\mathbf{H}$ or $\mathbf{A}$.

[2] See Appendix I.

[3] The Hamiltonian $\mathcal{H}$ and the action W must be considered as *functions*, not quantities. The variables in which they are expressed form a very essential part of their definition.

[4] R. Becker, "Theorie der Elektrizitat," *Elektronentheorie*, vol. II, Teubner, Leipzig, 1933, p. 69. This discovery was made by K. Schwarzschild in 1903.

[5] L. Brillouin, "A Theorem of Larmor of Importance for Electrons in Magnetic Fields," *Phys. Rev.*, vol. 67, pp. 260–266; April, 1945.

[6] The most general transformations which leave the field equations and the laws of motion invariant are the "gauge transformations." See W. Heitler, "The Quantum Theory of Radiation," Oxford, 1936, p. 3.

we can with impunity use the word "electron" for these indefinitely subdivided particles. We can now define invariants as quantities which are determined by the

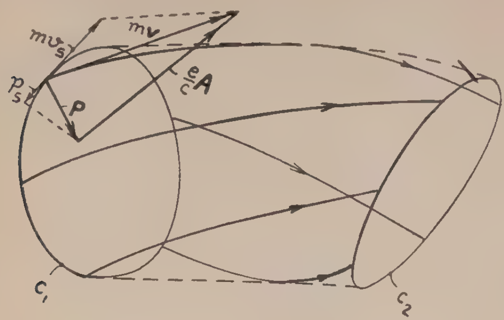


Fig. 1—Lagrange's invariant.

simultaneous state of a *group* of electrons, and *remain constant during their motion.* Invariants thus characterize certain particularly simple properties of an electron beam in a similar way as integrals characterize properties of single trajectories. Hence their importance and usefulness in the theory of electronic devices.

We consider a closed curve c_1 in an electron beam (Fig. 1) at some time t_1. The electrons which at the time t_1 are situated on c_1 will be situated at some later time t_2 on some other curve c_2. Let us determined in every point of the curve the component of the total momentum tangential to the curve, which we call p_s, and with the line element ds form the following integral round the closed circuit c

$$C = \int_c p_s ds.$$

This is called the "*circulation*" of the vector $\mathbf{p}$ around c. *It can be proved from Hamilton's canonical equations that*

$$C = \text{constant} \tag{5}$$

along all the positions successively occupied by the curve c. In particular, if $C=0$ at one position, it remains zero during the whole motion.[7]

By the well-known theorem of Stokes, the circulation can also be expressed as the flow of the vector curl $\mathbf{p}$ through *any* surface bounded by the curve c, hence the theorem (5) can also be expressed by saying that a closed circuit of electrons will carry the flux of the vector curl $\mathbf{p}$ with it, without any change. Let us apply this to an infinitesimal area dS, moving along with the electrons. Let n be the normal to dS and $\text{curl}_n\mathbf{p}$ the component of curl $\mathbf{p}$ in this direction. (n need not and will not in general coincide with the direction of motion). We now obtain the same theorem in the following differential form:

$$\text{curl}_n\mathbf{p}\, dS = \text{constant}. \tag{6}$$

This expression is Lagrange's "differential invariant," which he discovered in 1808. The form (5) is often called Poincaré's integral invariant,[8,9] as H. Poincaré derived it, in a more general form than the one here used, in 1890. In our special applications there is no need to use this name, as the equivalence of (5) and (6) was proved in 1845 by Sir George Stokes. Indeed, (5) is nothing else than the "circulation theorem" of hydrodynamics, slightly generalized, as the total momentum $\mathbf{p}$ for electrons has the value $m\mathbf{v}$, used in hydrodynamics, only in the absence of magnetic fields. As the circulation theorem of hydrodynamics was also first discovered by Lagrange, it appears only fair to call both (5) and (6) "Lagrange's Invariants."

[7] See Appendix II.

[8] E. T. Whittaker, "Analytical Dynamics," fourth edition, Cambridge, 1937, p. 267.

[9] C. Carathéodory, "Geometrische Optik," Berlin, Germany, 1937, p. 32.

Application of Lagrange's Theorem to Electron Beams

As a first application, let us show how an *integral of motion* can be derived from Lagrange's integral invariant in the practically important case of axially symmetrical electromagnetic fields (Fig. 2). In such fields the vector potential $\mathbf{A}$ is "tangential," that is, it runs in circles around the axis. It has the same direction as the velocity v_t in Fig. 2. If a group of electrons were situated at some time t_1 on a circle c_1, at some later time t_2 they will again occupy a circle, coaxial with the first. By reason of the rotational symmetry we can in this case immediately write down (5) in the following form:

$$(rv_t)_1 - (rv_t)_2 = \frac{e}{mc}\left[(Ar)_1 - (Ar)_2\right]$$

$$= \frac{e}{2\pi mc}(\Phi_1 - \Phi_2). \tag{7}$$

The second part of the equation follows immediately by applying Stokes' theorem to the circulation of the vector potential A. The curl of the vector $\mathbf{A}$ is the magnetic intensity $\mathbf{H}$, and its circulation is the magnetic flux Φ, hence we obtain the result that *the increase of angular momentum of an electron between two points of its trajectory is proportional to the difference of magnetic flux Φ through circles drawn through it coaxially, in its first*

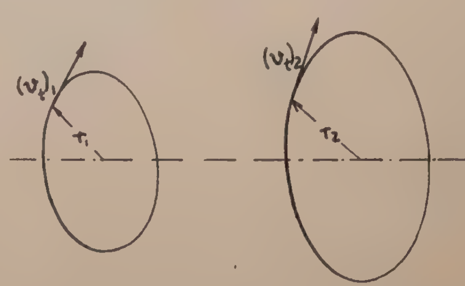


Fig. 2—Busch's theorem.

and in its second position. In the special case when the field is stationary, this flux difference is equal to the flux which flows through any surface bounded by the two circles. This is known as Busch's theorem, and is fundamental for electron optics.[10]

[10] For some further conclusions from this theorem, see D. Gabor, "Electron Optics," *Electronic Engineering*, December, 1942.

From the Hamiltonian analogy of dynamics and optics it must be expected that Lagrange's theorem has an optical counterpart. It will be useful, however, to recall to mind that an optical system can represent a dynamical one only under certain restrictions. First of all, as the idea of "time" is entirely alien to geometrical optics, only electron motion in steady fields can be represented optically.[11] A second qualification is that the kinetic energy must be a function of the position only; i.e., only "regular" beams, which have started from a cathode with zero velocity or from a point with constant velocity, can be represented, but not trajectories with arbitrary distribution of initial velocity. Finally, though magnetic fields can be represented by a refractive index, this is of a strange type which has no counterpart in ordinary optics. An electron in a magnetic field cannot describe the same trajectory forward and backwards, while light can always be sent back the same way, along the same ray.[12] If we accept these three restrictions, there exists a counterpart in optics of Lagrange's theorem. This is the theorem of Malus, discovered, curiously, in the same year as Lagrange's (1808).

According to Malus' theorem, if a system of rays starts at right angles to some surface, it is always possible to construct a family of surfaces which cut every ray at right angles. Such a system of rays is called in geometry and in optics a "*normal congruence.*" It follows from Lagrange's theorem that the same is true for electron trajectories in a purely electrostatic field, provided that they all start with zero or constant velocities from a point or at right angles to a surface. This is illustrated in Fig. 3.

On the other hand, Malus' theorem is *not* valid for electron trajectories in a magnetic field. An example is shown in Fig. 4. If the beam is imagined built up of different layers, it is possible to draw orthogonal curves to the trajectories on every one of these cigar-shaped surfaces, and connect these curves by surfaces, but these will not be at right angles to the trajectories, which form what is known as a "skew congruence."

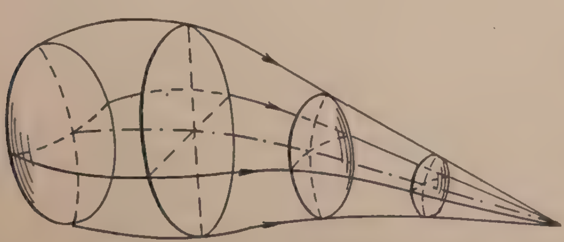


Fig. 3—Electron beam in electric field. "Normal congruence."

It may now be asked whether Malus' theorem cannot be extended also to electrons in a magnetic field, if instead of the velocity vector $\mathbf{v}$ we consider the vector of the total momentum $\mathbf{p}=m\mathbf{v}-(e/c)\mathbf{A}$, and try to draw surfaces at right angles to these. This is indeed possible in many cases, though not in all. The problem

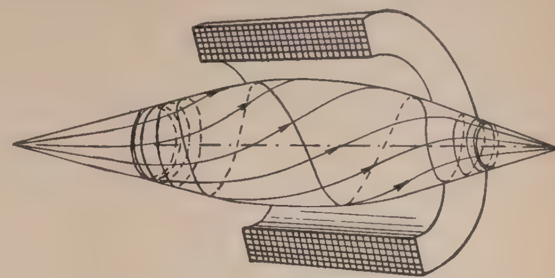


Fig. 4—Electron beam in magnetic field. "Skew congruence."

which is of particular importance to us in electronics is the representation of electron beams emitted by cathodes at zero velocity ("regular beams"). The results are as follows:[13]

If the magnetic field has no component normal to the cathode, the vector $\mathbf{p}$ will form in the electron beam an *irrotational field;* i.e., we have everywhere

$$\operatorname{curl} \mathbf{p} = 0. \tag{8}$$

In this case we shall be able to treat electron motion in an electromagnetic field, stationary or not, by the same methods which apply to electric fields, merely by formulating the equations with the help of the total momentum $\mathbf{p}$ instead of $m\mathbf{v}$.

If, however, the magnetic field at the cathode has a normal component, we have at the cathode surface

$$\operatorname{curl}_n \mathbf{p} = -\frac{e}{c} H_n \tag{9}$$

and by Lagrange's theorem the field of $\mathbf{p}$ in the beam cannot be irrotational. This is a much more complicated case, and we shall leave it aside in this paper. Fortunately, in almost all electronic devices (8) is fulfilled, and those in which it is not fulfilled (e.g., the Farnsworth dissector, the Orthicon, the Phillips image converter) can be satisfactorily discussed by elementary methods. Perhaps the only exception is the magnetron with skew magnetic field, electron dynamically the most complicated of all devices. This could be treated only by an extension of the simple theory discussed in this paper.

The Hamilton-Jacobi Equation

Let us now consider beams in which (8) applies; i.e., in which $\mathbf{p}$ is an irrotational vector. This suggests naturally (following the example of electromagnetic theory, or of hydrodynamics) to express $\mathbf{p}$ as the derivative of some potential function which may be called W. We write, therefore

$$\mathbf{p} = \operatorname{grad} W \tag{10}$$

and with this we have solved (8). There remains the

[11] In three dimensions. A more general optical analogy exists in four dimensions. See D. Gabor, "Electron Optics," *Electronic Engineering,* February, 1943, and also later on in this paper.

[12] Dr. Alfred N. Goldsmith has pointed out to me that this is true only in the absence of magneto-optical effects.

[13] See Appendix III.

problem of determining the function W in such a way that it shall solve at the same time the equations of motion; i.e., give the correct value of $d\mathbf{p}/dt$. This problem was solved by Hamilton in 1834, in the following way: To (10), which, reduced to the components, can be written

$$\frac{\partial W}{\partial x} = p_x \quad \frac{\partial W}{\partial y} = p_y \quad \frac{\partial W}{\partial z} = p_z \tag{10a}$$

we add a further equation

$$\frac{\partial W}{\partial t} = -\mathcal{H}^*(x, y, z, t). \tag{11}$$

The function $\mathcal{H}^*$ is again the total energy, but unlike the Hamiltonian $\mathcal{H}$ it is a function of position and time only, as p_x, p_y, p_z have been expressed in it as functions of x, y, z, and t. $\mathcal{H}^*$ has therefore a more restricted meaning than $\mathcal{H}$. While the Hamiltonian characterizes a *dynamical problem*, $\mathcal{H}^*$ characterizes a *congruence of trajectories*, which we have called a "regular" electron beam.

The time derivative $\delta W/\delta t$ cannot be arbitrarily chosen, as the space derivatives are given by (10a), and three compatibility equations of the type $\delta^2 W/\delta x\delta t = \delta^2 W/\delta t\delta x$ must be fulfilled between them. But these turn out to be the canonical equations of motion; hence (11) is indeed a valid formulation of the dynamical problem.

We now write out (11) in full, substituting the values of p_x, etc., from (10a), and obtain

$$\left(\frac{\partial W}{\partial x} + \frac{e}{c} A_x\right)^2 + \left(\frac{\partial W}{\partial y} + \frac{e}{c} A_y\right)^2 + \left(\frac{\partial W}{\partial z} + \frac{e}{c} A_z\right)^2 = 2m\left(e\phi - \frac{\partial W}{\partial t}\right). \tag{12}$$

This is the nonrelativistic Hamilton-Jacobi[14] equation for an electron in an electromagnetic field.

The derivation which has been sketched out here, and the use which we are going to make of this equation are rather different from the usual practice in analytical dynamics.[15,16] There it is shown that, once a *general* solution of the Hamilton-Jacobi equation is known, *any* trajectory can be derived from it. But this is of little use in electronics. The Hamilton-Jacobi equation can be exactly and generally integrated only in a few simple cases, in which the solutions can be obtained also by elementary methods. On the other hand *particular* solutions with the correct initial conditions often can be fairly easily obtained by numerical or graphical methods. Therefore we use the Hamilton-Jacobi equation only to determine an "*action function*" $W(x, y, z, t)$, which is a sort of "velocity potential[17-19]" for electron beams with irrotational momentum distribution. As we have seen, this application is possible only in the case of beams which are emitted by a cathode without a magnetic component normal to its surface. In the case of the "skew magnetron" the Hamilton-Jacobi equation can also be applied, but only in the orthodox manner, which is not in general very helpful.

Application of the Hamilton-Jacobi Equation

The least promising line of attack on the Hamilton-Jacobi equation is a direct attempt to integrate it analytically. On the other hand, this equation may be very useful in three types of applications:

(*a*) approximate solutions
(*b*) the inverse dynamical problem
(*c*) small variations of dynamical problems.

We restrict ourselves for a start to steady or quasi-steady fields, and illustrate the applications by two-dimensional examples.

(*a*) *Graphical or Numerical Approximations*

The Hamilton-Jacobi equation in the case of a pure electrostatic field

$$\left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 = (\text{grad } W)^2 = 2me\phi \tag{13}$$

admits a very simple interpretation which leads to a useful method of graphical solution. Fig. 5 is an illustration in the case of plane motion. Let us assume that we know one line W=constant. We can now draw the next line $W = W_0 + dW$, if we write (13) in the form

$$(dW/dn)^2 = 2me\phi.$$

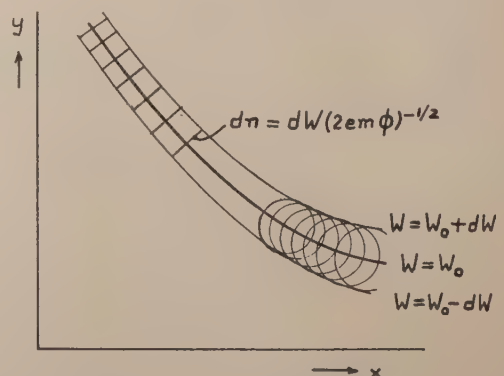


Fig. 5—Propagation of the action in an electrostatic field.

[14] This equation was found by Hamilton in 1834, by Jacobi only in 1837, but the customary name "Hamilton-Jacobi equation" is a well-deserved recognition of Jacobi's merits in the development of Hamiltonian dynamics.

[15] See p. 315 of footnote reference 8.

[16] A. Sommerfeld, "Atomic Structure and Spectra," Appendix VII.

[17] A velocity potential appears to have been first used in electron optics by F. Gray, "Electrostatic Electron Optics," *Bell Sys. Tech. Jour.*, vol. 18, pp. 1–32; January, 1939, where it is used for electrostatic fields only, though extension to magnetic fields is also mentioned.

[18] The Hamilton-Jacobi equation has been also used by K. Spangenberg in a paper, "Use of the action function to obtain the general differential equations of space charge flow in more than one dimension." *Jour. Frank. Inst.*, vol. 232, p. 365; October, 1941.

[19] K. Spangenberg and L. M. Field, "Some simplified methods of determining the optical characteristics of electron lenses," Proc. I.R.E., vol. 30, pp. 138–144; March, 1942; and *Electrical Communication*, vol. 20, p. 305; April, 1942. Both Gray and Spangenberg consider only steady fields.

This means that we must proceed by a distance $dn = \pm dW/\sqrt{2me\phi}$ in the direction of the normal. The ambiguity of sign means only that we can assume the action increasing in either of the two directions; i.e., the electrons can be assumed to move one way as well as the other. Once we have fixed the sense of motion, we can gradually construct by this process the whole family of lines of constant action, and draw the trajectories as the curves which intersect them orthogonally. This can be done with fair accuracy if a mirror is used as a ruler, which is turned until the kink disappears between the line and its image.

Instead of drawing the lines dn at right angles to W=constant, we can also use Huygens' construction; i.e., draw circles with the radius dn and construct the next line of constant action as their envelope. It is known that this formal analogy with the construction of a wave front has received a very concrete physical interpretation in wave mechanics.[20]

In the static case *with magnetic field* the Hamilton-Jacobi equation becomes

$$\left(\frac{\partial W}{\partial x} + \frac{e}{c} A_x\right)^2 + \left(\frac{\partial W}{\partial y} + \frac{e}{c} A_y\right)^2 = \left(\text{grad } W + \frac{e}{c} \mathbf{A}\right)^2 = 2me\phi. \qquad (14)$$

It has been mentioned that the vector potential is not entirely specified, as we can always add the gradient of an arbitrary scalar function to it. We now use this to make **A** zero at the cathode surface, so that the cathode itself shall become a surface $W=0$, and in the immediate neighborhood of the cathode W shall take the same course as in the purely electrostatic case. For example, in the case of a magnetron with cathode radius r_0 a vector potential in tangential direction

$$A_t = \tfrac{1}{2} H(r - r_0^2/r)$$

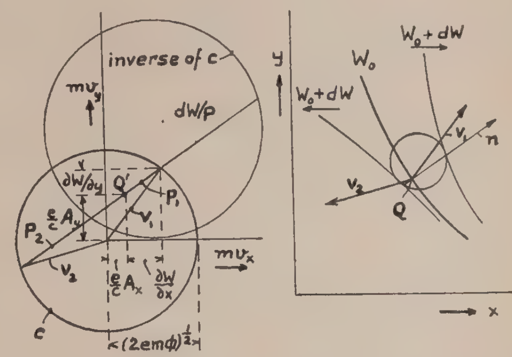
fulfills the conditions curl $\mathbf{A} = \mathbf{H}$ and $\mathbf{A} = 0$ for $r = r_0$.

Once the vector potential is thus suitably specified, the same process can be applied to solve (14) as in the electrostatic case, but now it is convenient to draw two figures, as shown in Fig. 6. In the figures at the left the co-ordinates are the mechanical momenta

$$mv_x = \frac{\partial W}{\partial x} + \frac{e}{c} A_x \qquad mv_y = \frac{\partial W}{\partial y} + \frac{e}{c} A_y$$

and in these co-ordinates the Hamilton-Jacobi equation is represented by a circle C with radius $\sqrt{2me\phi}$. In the diagram at the right the co-ordinates are x and y. Again we assume that in this "configuration" space we know one line (or surface) $W = W_0$. To a point Q we mark a corresponding point Q' in "momentum space," with the co-ordinates $(e/c)A_x$ and $(e/c)A_y$, and through this point Q' we draw a line parallel to the direction of the normal n. This intersects the circle in two points, and gives two velocities, v_1 and v_2, equal in magnitude but different in direction, pointing towards opposite sides of the line $W = W_0$. The difference in direction illustrates once more the fundamental property of the magnetic field, that a trajectory cannot be described in opposite directions.

In order to find the distances dn we must now take the reciprocals of the distances $p_1 = (dW/dn)_1$ and $p_2 = (dW/dn)_2$, and multiply by dW. It is more instructive, however, to leave the direction of **p** for the moment undetermined, and determine the locus of the radii dn. This means that we must transform the circle c by "reciprocal radii" or "inversion" with respect to the point Q'. But the inverse of a circle is again a circle, hence the locus of $dn = dW/p$ is easily determined. This inverse circle, transferred into the second diagram is an elementary "Huygens' wavelet." Its eccentricity with respect to Q illustrates the peculiar effect of the magnetic field, which allows the "action waves" to spread more easily in some directions than in others.



Momentum space Configuration space

Fig. 6—Propagation of the action in an electromagnetic field.

In practical constructions it is not necessary to draw the Huygens' wavelets in every point, and the process can be made fairly speedy by drawing, once for all, two reciprocal figures, say a straight line and its inverse circle, so that reciprocal distances can be drawn without recourse to the slide rule.

(b) *The Inverse Dynamical Problem*

The electronic engineer is often faced not with the problem of finding electron motion in a given field, but designing the field so as to produce a certain desired type of electron motion. The Hamilton-Jacobi equation is particularly suitable for this purpose, as once the motion is given by some action function $W(x, y, z, t)$ the potential ϕ which will produce this motion is immediately given by

$$2me\phi = 2m \frac{\partial W}{\partial t} + (\text{grad } W)^2. \qquad (15)$$

In the following we will restrict ourselves to steady motion. In this case, what the designer wants to prescribe are usually only the trajectories, the velocities

[20] See the extremely lucid explanations of E. Schrödinger in his second paper on "Wave Mechanics," reprinted in his "Collected Papers on Wave Mechanics," Blackie, 1928, which we have followed up to this point.

themselves are mostly not of immediate interest. All one has to do is to draw the family of trajectories which form the beam, and construct graphically or analytically the family of curves which intersect these at right angles; i.e., the lines W=constant. Let their equation be $u(x, y)$=constant. We can now make W any arbitrary function of u, and obtain the potential which can produce the prescribed trajectories in a form

$$\phi = F(u)(\text{grad } u)^2 \tag{16}$$

where F is an arbitrary but positive function of u. This can be modified as it suits the purpose best.

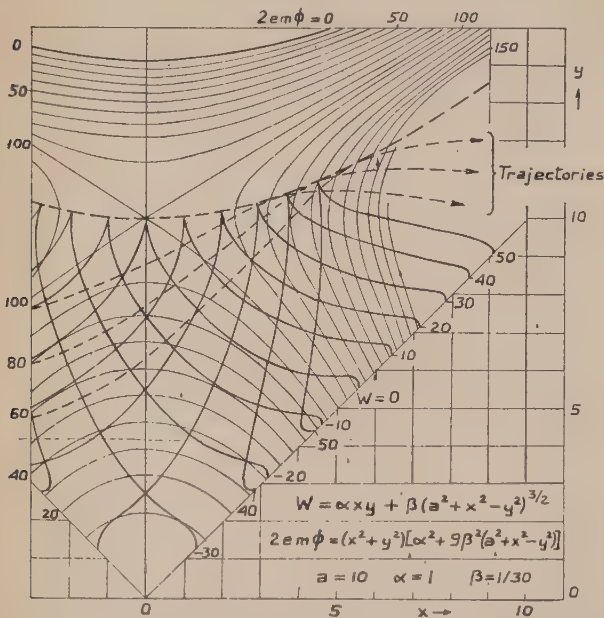


Fig. 7—Example of two-valued action function. Equipotential lines are drawn with thin continuous lines, the loci of constant action with thick lines.

The action W cannot be arbitrarily prescribed in cases in which W is a multiple-valued function, as this would result in general in a multiple-valued potential, which is an impossibility. Multiple-valued action functions arise in all problems in which electron paths cross one another.[21] In this case the following artifice may be adopted:

Let us assume two orthogonal families of curves

$$u(x, y) = C_1 \qquad w(x, y) = C_2.$$

If we want a two-valued action function, we can assume it in the form

$$W = f(u) + [g(w)]^{1/2} \tag{17}$$

where f and g are arbitrary but single-valued functions. The potential ϕ follows now from the equation

$$2me\phi = \left(\frac{df}{du}\right)^2 (\text{grad } u) + \frac{1}{4g}\left(\frac{dg}{dw}\right)^2 (\text{grad } w)^2$$

which is a single-valued function. The term containing $\sqrt{g(w)}$ has dropped out because of the orthogonality of u and w. If four-valued action functions are required, we can put

$$W = [f(u)]^{1/2} + [g(w)]^{1/2}.$$

A certain class of orthogonal curves can be easily manufactured, by assuming u and w as conjugate harmonic functions. This is to say we assume an arbitrary function F of the complex variable $z=x+jy$, and put

$$F(x + jy) = u(x, y) + jw(x, y).$$

As an example, let us put

$$F(z) = z^2 = x^2 - y^2 + 2jxy.$$

We choose, for instance,

$$W = \alpha xy + \beta(a^2 + x^2 - y^2)^{3/2}$$

which gives

$$2em\phi = \text{grad}^2 W = (x^2 + y^2)\left[\alpha^2 + 9\beta^2(a^2 + x^2 - y^2)\right].$$

A numerical example of this field is shown in Fig. 7. The lines W=constant have two branches each, which terminate in a common cusp at the hyperbola

$$y^2 - x^2 = a^2$$

beyond which the action becomes complex. This hyperbola is the envelope of the trajectories belonging to the beam.

Though it is easy in this way to manufacture electron beams together with their corresponding potentials, these potentials cannot always be realized. They will not in general satisfy the Laplace equation (in the case of negligible space charge), or the Poisson equation (in the case of strong electron currents). Nevertheless, they may often give useful hints to the designer; and in many cases it will be possible, if necessary, to eliminate the space charge, or space-charge deficiency, by superimposing a relatively weak field, and treating this by the perturbation methods discussed in the next section.[22]

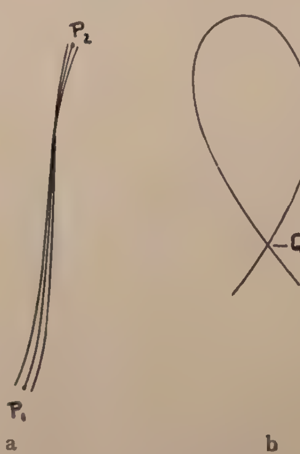


Fig. 8—Data of motion which determine the electrostatic field.

How much of the electron motion are we allowed to prescribe arbitrarily in a Laplacian field? Fig. 8 gives an answer to this, in the case of plane motion. It can be

[21] See J. H. L. Jonker and A. J. W. M. van Overbeek, "A new converter valve," *Wireless Eng.*, vol. 15, p. 423; August, 1938, which contains some instructive examples of such electron beams, illustrated by means of a rubber membrane model.

[22] A direct attack on the problem of electron motion consistent with its own space charges leads to equations of formidable complication and difficulty. See footnote reference 19.

proved that the electrostatic field will be entirely determined by a "narrow beam" between two points P_1 and P_2, if the potentials in two of its points, e.g., in the end points, are prescribed (Fig. 8a). This is analogous to the well-known theorem for paraxial beams in axially symmetrical fields.

An important exception must be noted. If the central trajectory forms a closed loop; i.e., if it intersects itself, and there is no electrode inside this loop, the field will be entirely determined by the loop of the trajectory and the potential of one of its points, say the point Q in which it intersects itself (Fig. 8b). All the rest of the field and of the beam can be constructed from these data.

(c) *Small Perturbations*

It is very often not sufficient for the electronic engineer to know the electron motion in a given field; in many cases he will be as much, or even more, interested by the changes in the motion produced by changes in the electromagnetic field. Perhaps the greatest advantage of the Hamilton-Jacobi equation is that it lends itself so well to the treatment of slight variations of a dynamical problem. The method of perturbations was used in celestial mechanics for almost 80 years before 1916, in which year the astronomers Epstein and Schwarzschild introduced the Hamilton-Jacobi methods into atomic physics.

Let us assume that we have found a solution of the Hamilton-Jacobi equation in a potential field ϕ_0, which we call W_0. We have therefore

$$\left(\frac{\partial W_0}{\partial x}\right)^2 + \left(\frac{\partial W_0}{\partial y}\right)^2 = 2em\phi_0.$$

We now add to ϕ_0 a small "perturbing" potential ϕ_1. Neglecting the square of the momenta produced by the perturbation we obtain the *linear* equation for W_1

$$\frac{\partial W_0}{\partial x}\frac{\partial W_1}{\partial x} + \frac{\partial W_0}{\partial y}\frac{\partial W_1}{\partial y} = em\phi_1. \tag{18}$$

It is now convenient to introduce new co-ordinates instead of x and y. We could, for instance, choose the unperturbed trajectories themselves and the family of curves orthogonal to them as suitable curvilinear co-ordinates. But there is no need to specify this more exactly, as the co-ordinate at right angles to the unperturbed trajectory drops out of (18) and one obtains the very simple result

$$\frac{\partial W_1}{\partial W_0} = \frac{\phi_1}{2\phi_0}. \tag{19}$$

We need only integrate this along an unperturbed trajectory to find the changes which the perturbing field has caused in the action function.

To illustrate this very useful method, let us consider the deflection of an electron stream in a triode by a positive grid (Fig. 9). As a convenient "unperturbed state" we choose the state in which the grid is uncharged, so that the trajectories are parallel, straight lines. If now we apply to the grid potentials slightly smaller than in the "unperturbed" state, the lines W=constant will move a little towards the anode, and curve slightly towards it. Their shape can be determined by an easy integration. If the grid charge is slightly positive, the effect is the opposite, the constant action lines move towards the cathode and curve away from the cathode. Consequently, the electron beams, which at negative grid charge were slightly concentrated towards the anode, now spread out.

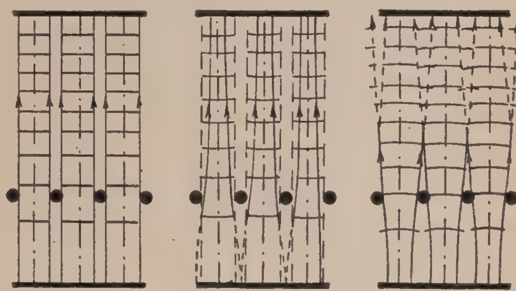


Fig. 9—An application of the perturbation method.

It can be seen from the figure that the action which at uncharged grid was a single-valued function now splits up into several branches. The lines W=constant intersect above the center of the grid wires. These parts of the W-lines and of the trajectories must be determined by extrapolation. This is not difficult, nor very arbitrary so long as the perturbation remains small. But even large departures of the grid potential from the uncharged state can be treated in this way, if the change is divided up into sufficiently small steps, which are calculated in succession, each step as the perturbation of the previous one.

Electron Motion in Rapidly Varying Fields (Transit-Time Effects)

If the electromagnetic field varies appreciably during the time of flight of an electron in the device, the energy integral is no longer valid and elementary theory loses its chief guide. But the Hamilton-Jacobi equation

$$\left(\frac{\partial W}{\partial x} + \frac{e}{c}A_x\right)^2 + \left(\frac{\partial W}{\partial y} + \frac{e}{c}A_y\right)^2 + \left(\frac{\partial W}{\partial z} + \frac{e}{c}A_z\right)^2 = 2m\left(e\phi - \frac{\partial W}{\partial t}\right) \tag{12}$$

remains valid, and the term $\partial W/\partial t$ indicates directly the departure from the constancy of total energy.

It appears desirable to treat time on the same footing as the spatial co-ordinates, but in (12) these figure in a markedly asymmetrical manner. It is known that relativity theory establishes a certain symmetry between time and space, let us therefore first see the relativistic Hamilton-Jacobi equation[16,20]

$$\left(\frac{1}{c}\frac{\partial W}{\partial t} - mc - \frac{e}{c}\phi\right)^2 - \left(\frac{\partial W}{\partial x} + \frac{e}{c}A_x\right)^2 - \left(\frac{\partial W}{\partial y} + \frac{e}{c}A_y\right)^2 - \left(\frac{\partial W}{\partial z} + \frac{e}{c}A_z\right)^2 = (mc)^2. \quad (20)$$

This, however, is not symmetrical, but rather antisymmetrical in t and x, y, z which is, of course, to be expected, as in relativity not the real time t, but the imaginary quantity jct is a counterpart of the spatial co-ordinates.

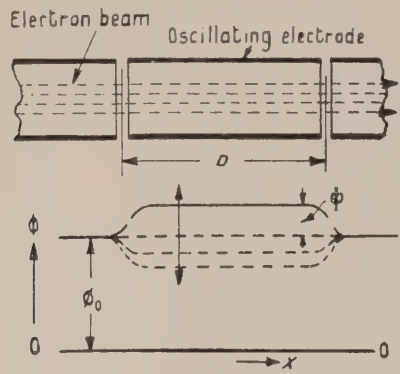


Fig. 10

In electronic devices we deal mostly with electron velocities considerably smaller than the velocity of light, hence the difference between (12) and (20), which consists of terms of the order $(v/c)^2$, is usually negligible. There is, therefore, no objection against adding other terms of this order to (12) by which the term at its right side is transformed into a negative square, instead of a positive square. Nor is it necessary to make the quantity c equal to the velocity of light, we can make it any velocity C, provided it is large against v. We introduce now as fourth co-ordinate

$$\tau = Ct \quad (21)$$

and obtain the Hamilton-Jacobi equation in the symmetrical form

$$\left(\frac{\partial W}{\partial \tau} + mC\right)^2 + \left(\frac{\partial W}{\partial x} + \frac{e}{c}A_x\right)^2 + \left(\frac{\partial W}{\partial y} + \frac{e}{c}A_y\right)^2 + \left(\frac{\partial W}{\partial z} + \frac{e}{c}A_z\right)^2 = 2m(e\phi + \tfrac{1}{2}mC^2). \quad (22)$$

We call this the *"antirelativistic"* Hamilton-Jacobi equation. If C tends to the limit infinity it goes over into the nonrelativistic equation, and in general the errors will be of the order

$$v^2\left(\frac{1}{c^2} + \frac{1}{C^2}\right).$$

The antirelativistic equation[23] can be interpreted as follows:

We can replace a time-dependent problem of electron motion in n dimensions by a stationary problem in n+1 dimensions, if we inpart to the electrons an initial velocity in the direction of the new co-ordinate large against the velocities in the other directions. In other words, the trajectories must always include small angles with the new axis.

This principle has been used for making a mechanical *static* model of a velocity-modulation tube of the Hahn-Metcalf type.[24,25] These tubes contain a "modulator," a section of which is shown in Fig. 10, and a "demodulator" of identical design. The electron beam emitted by a cathode is accelerated by a direct-current field, outside the figure, at the left, and moves in turn through a tube at constant potential ϕ_0, through an oscillating electrode, and through a third tubular electrode, called the "drift tube," which has the same constant potential ϕ_0 as the first. The beam is subjected to electric fields only in the two "gaps" at the entrance and at the exit of the oscillating electrode. The potential profile at different instants is indicated in the lower part of Fig. 10.

In the mechanical model, a photograph of which is shown in Fig. 11, the potential distribution is represented in the form of a relief. One horizontal co-ordinate x represents the axial position, the other co-ordinate t represents time, the vertical co-ordinate is proportional to ϕ. Any section at t = constant represents the potential profile in the tube at that instant. In addition to the modulator shown in Fig. 10, the relief in Fig. 11 comprises also at the left a constant downward slope, which represents the accelerating direct-current field, and at the right a constant rising slope, representing a retarding direct-current field. This has been added in order to measure the energy acquired by the electrons by the highest point which they reach on this slope.

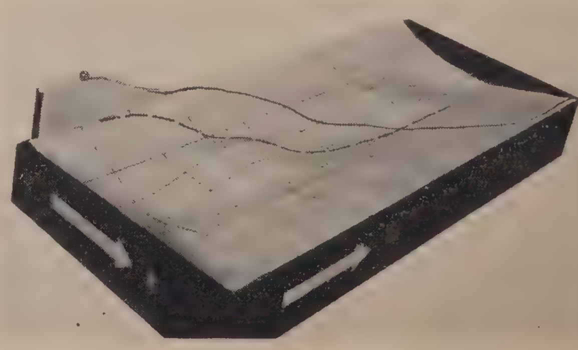


Fig. 11—Static model of Hahn-Metcalf velocity-modulation tube.

The left edge of the model represents the cathode. Electrons, represented by steel balls, leave this with

[23] I have called this an antirelativistic equation, but it is, of course, one of the minor fruits of relativity. So long as classical Newtonian mechanics was believed to have absolute validity, mathematicians spent most of their effort in finding rigorous solutions of the classical equations. But as relativity has taught us that the classical laws are themselves only approximations, there is no reason why we should not replace them by more convenient approximations.

[24] D. Gabor, "Energy conversion in electronic devices," *Jour. I.E.E.* (London), part III, vol. 91, p. 128; September, 1944.

[25] W. C. Hahn and G. F. Metcalf, "Velocity-modulated tubes," Proc. I.R.E., vol. 27, pp. 106–116; February, 1939.

zero velocity; that is to say, their $x(t)$ curves start here in the direction of time. According to the principle mentioned, the "velocity in the time direction" must be large as compared with the physical velocities; i.e., the steel balls must be launched from a height large as compared with the height differences in the relief. In the model, which is intended for demonstration only, not for measurements, this rule has not been strictly observed, in order to display the trajectories on a more convenient scale.

If a steel ball is launched in a certain "phase," corresponding to the dotted line, the trajectory turns parallel to the time axis at the exit from the modulator, which means that the corresponding electron has lost all its energy in the modulator. A steel ball launched half a cycle later—continuous line—climbs on the slope at the right to a height above the cathode edge; i.e., it has gained energy from the alternating field in the modulator. The two trajectories intersect at a certain point. At this point the electron which was launched

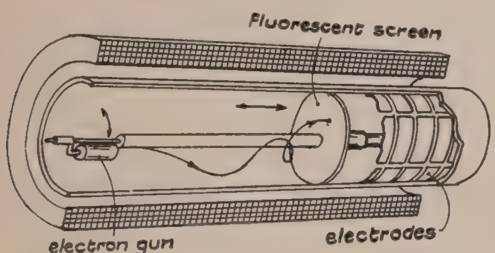


Fig. 12—Static model of oscillating magnetron.

later has caught up with the one launched half a cycle earlier. This phenomenon is called " phase focussing" or "bunching."

Mechanical models of this sort can be constructed only for one-dimensional transit-time problems, but another kind of model is possible also for two-dimensional motion. Fig. 12 is a suggested *static* model of an oscillating magnetron. The electrodes are longitudinally subdivided and direct-current potentials are impressed on them, which vary along the axis by the same law as they would vary with time in actual operation. An electron gun shoots a thin beam with large velocity in axial direction, tangentially to the electrode which represents the cathode. The trajectory is registered on a fluorescent screen which can slide axially along the cathode.

The new principle suggests also another kind of model, in which electron trajectories are replaced by light rays. Hitherto, it was not thought possible to imitate electron optical devices optically, as the enormous ratios of electron-optical refractive indexes could not be reproduced. But if we add a new dimension to the device, only small differences of the refractive index will be required to produce appreciable deflections. Instead of trying to produce a continuously varying refractive index, we can also produce these deflections by suitably shaped refracting surfaces. Fig. 13 shows an optical model of the same velocity-modulation tubes of which we have already seen a mechanical model. The surfaces of the thin refracting plates of glass or transparent plastics must be shaped according to the po-

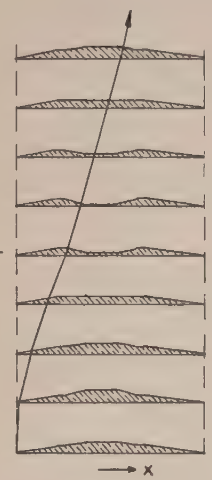


Fig. 13—Optical model of velocity-modulation tube.

tential profile. Unlike the mechanical model, this one is suitable also for two-dimensional problems, and it has the advantage of being free from friction.

Perhaps more important than these new models is the discovery that there exists a fairly highly developed theory of transit-time phenomena in a rather unexpected place; in the Gaussian theory of optical instruments. Gaussian optics is the theory of light rays which include small angles with an axis. The special case of rotation-

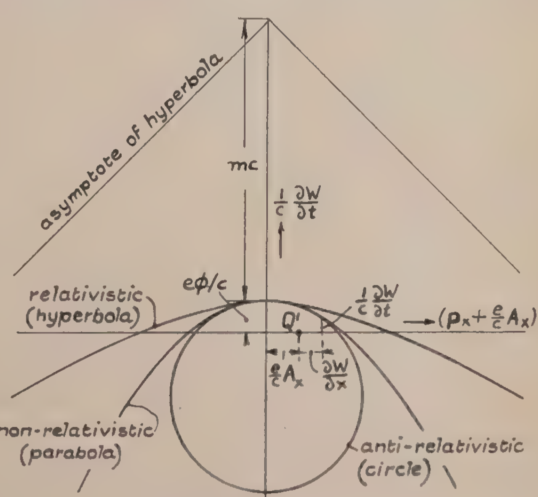


Fig. 14—The three types of Hamilton-Jacobi equation.

ally symmetrical optical instruments, which is most widely known, is of little use in electronics, but the general case also has been worked out in considerable detail, especially by Gullstrand, and by Carathéodory.[26]

The analytical and graphical methods explained in the previous chapter can now be transferred to transit-time problems, with the help of the "antirelativistic" Hamilton-Jacobi equation. Fig. 14 explains the procedure

[26] See chapter V, pp. 84–102, of footnote reference 9.

in the case of one-dimensional motion, which is entirely analogous to the treatment of two-dimensional stationary problems. It illustrates particularly the difference between the exact, relativistic Hamilton-Jacobi equation, and the two approximations. The locus of the total momenta in relativistic treatment is a hyperbola, in the classical treatment a parabola, in the "anti-relativistic" approximation a circle (in the case of two- or three-dimensional motion a sphere or a hypersphere). The corresponding inverse figures; i.e., the "Huygens' wavelets," are shown in Fig. 15. So long as C is assumed very large (as in the drawings) the three methods will give practically the same results for small velocities.

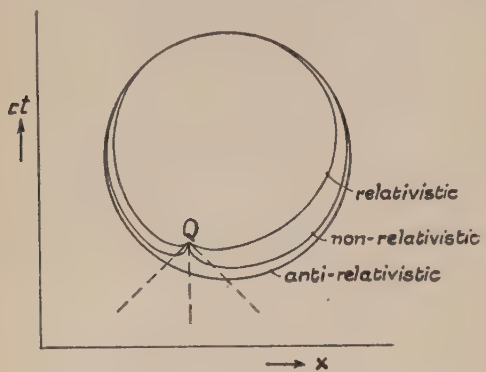


Fig. 15—The three types of Huygens' wavelets.

Electron Beams with Random Motion

We add now a further touch of realism to the abstractions hitherto considered, by taking into consideration the random distribution of the initial velocities in direction and magnitude. This means that we must now drop the picture of the electron beam as a fluid in three dimensions, which can have only one (or a finite number) of velocities in any given point. Fortunately, we can immediately replace it by another picture, by the flow of a fluid in Hamilton's *phase space.*

Hamilton's idea of the phase space composed of configuration space and momentum space is one of the most useful of his creations.[27] In the special case of electron motion this space is six-dimensional for three-dimensional problems; hence only one-dimensional problems can be graphically illustrated, but these give sufficient help to the imagination to deal also with more complicated cases.

The usefulness of phase space is based on the second general invariant of dynamics, Liouville's invariant, which may be stated without proof.[28] The motion of the points representing electrons in phase space is like *the motion of an incompressible fluid.* If these points, or let us simply say electrons (though we are now at two removes from the original idea of the electron as a definite point charge), occupy at some instant a closed surface in phase space, the volume of this space will remain invariant during the motion. As the number of electrons remains also invariant, we can express this by saying that the density D in phase space is an invariant of the motion. This is illustrated in Fig. 16.

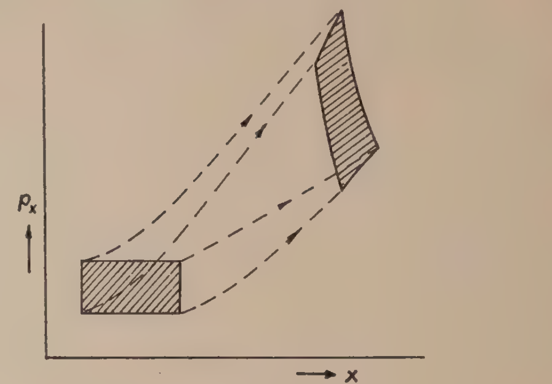


Fig. 16—Liouville's invariant.

The "continuity equation" in phase space is therefore

$$\frac{dD}{dt} = 0. \tag{23}$$

To simplify matters, we will write this out in full only for the case of one-dimensional motion, writing x for the co-ordinate, and p for the corresponding momentum

$$\frac{dD}{dt} = \frac{\partial D}{\partial t} + \frac{dx}{dt}\frac{\partial D}{\partial x} + \frac{dp}{dt}\frac{\partial D}{\partial p} = 0. \tag{24}$$

We now substitute dx/dt and dp/dt from the canonical equations (2), and obtain

$$\frac{dD}{dt} = \frac{\partial D}{\partial t} + \frac{\partial \mathcal{H}}{\partial p}\frac{\partial D}{\partial x} - \frac{\partial \mathcal{H}}{\partial x}\frac{\partial D}{\partial p} = 0. \tag{25}$$

This is the fundamental equation for the electron distribution. Its generalization to more than one dimension is obvious.

Electron Distribution in Stationary Electromagnetic Fields

If the field is stationary, the motion will also have a stationary solution and the corresponding distribution density D will be the solution of the equation

$$\frac{\partial \mathcal{H}}{\partial p}\frac{\partial D}{\partial x} - \frac{\partial \mathcal{H}}{\partial x}\frac{\partial D}{\partial p} = 0. \tag{26}$$

This is a homogeneous linear partial differential equation of the first order, with the general solution

$$D = F(\mathcal{H}). \tag{27}$$

As in stationary fields $\mathcal{H}$ = constant along every trajectory, this means that *the phase density D is a constant along every trajectory*, though it can vary from one trajectory to another in any arbitrary way. The validity

[27] It might be argued that Hamilton was only the grandfather of the phase space, and that it was J. Willard Gibbs who recognized its significance and usefulness. See Gibbs' "Elementary Principles of Statistical Mechanics," 1902. The name "phase space" is also due to Willard Gibbs.

[28] For a simple proof, see R. H. Fowler, "Statistical Mechanics," Cambridge, 1929, p. 11. For more rigorous proofs, see p. 283 of footnote reference 8, and p. 39 of footnote reference 9.

of this result is not confined to the one-dimensional example.

Let us now apply this to an electron beam emitted by a thermionic cathode. For simplicity let us assume that the cathode emits uniformly over its emitting area, where the density is

$$D = K \exp\left(-\frac{mv^2}{2kT}\right). \tag{28}$$

From this it follows at once that any point of the phase space we shall have either the density

$$D = K \exp\left(-\frac{1}{kT}\left(\tfrac{1}{2}mv^2 - e\phi\right)\right) \tag{29}$$

or the density

$$D = 0 \tag{30}$$

according to whether this point of phase space is reached by electrons or not. The whole solution of the distribution problem can therefore be put together out of the "Maxwell-Boltzmann distribution" (29), and the trivial solution (30).

As the two solutions are known, the only problem which remains is *to fix the boundary between them.* But this is a dynamical problem of the type as considered in the previous sections; hence the problem of random motion is effectively reduced to one of a simpler type.

As an example, let us first consider one-dimensional motion, between plane electrodes. As is evident from their derivation, the solutions (29) and (30) are not affected by the presence of a magnetic field, though the boundary between them may be affected, but in the one-dimensional case a magnetic field (which must be in the direction of the motion if the motion is to remain one-

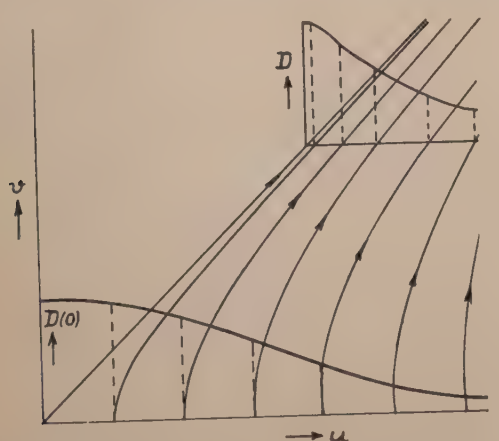


Fig. 17—Velocity distribution in the case of linear motion.

dimensional) has no effect at all. We can therefore neglect it from the start. It is now convenient to express the momentum p by the velocity, but we now call this u, to distinguish it from the velocity $v=\sqrt{2e\phi/m}$ which a "regular" beam; i.e., one without initial velocity, would have at the same point. We can now write the Hamiltonian

$$\mathcal{H} = \tfrac{1}{2}mv^2 - e\phi = \tfrac{1}{2}m(u^2 - v^2)$$

hence the density becomes

$$D = D(u^2 - v^2).$$

The density remains constant along the hyperbolas $u^2 - v^2 =$ constant which are, of course, the lines of constant total energy. This and the apparent deformation of the density distribution are shown in Fig. 17. The boundary between the Maxwell-Boltzmann solution and the trivial solution $D=0$ is very simple in this case, as it is represented by the line $u=v$. This means that only those electrons can appear in the *beam* which have left the cathode, $\phi=0$ with zero or positive velocity.

In the case of space-charge-limited flow, ϕ has to be measured from the potential minimum, not from the cathode surface.

In problens of more than one dimension the determination of the boundary is somewhat more complicated and we cannot go into it in detail. Fig. 18 gives a hint how this can be carried out. The solid angle outside which no electrons can reach a point P is determined by those electrons which have started from the edge of the cathode, with initial velocities tangential to it. As in most cases, this solid angle is very small, approximate solutions can easily be obtained by similar methods as discussed in the previous chapters.

Electron Distribution in Rapidly Varying Fields

In transit-time problems, the energy integral $\mathcal{H} =$ constant and the simple law $D = D(\mathcal{H})$ fail us simultaneously, and they cannot be replaced by other, more general solutions. Nevertheless, as the equation for D is a linear partial-differential equation of the first order, solutions in special cases can be developed fairly easily. These may be of some interest for the problem of the theoretical limitation of the performance of electronic devices by random motion.[29,30]

Fig. 18—Angular limitation of an electron beam.

Conclusion

We have seen that certain methods of dynamics originating from Hamilton's work have appreciable advantages in the field of electronic problems discussed

[29] D. B. Langmuir, "Theoretical limitations of cathode-ray tubes," Proc. I.R.E., vol. 25, pp. 977–991; August, 1937.
[30] J. R. Pierce, "Theoretical limitations to transconductance in certain types of vacuum tubes," Proc. I. R. E., vol. 31, pp. 657–663; December, 1943.

in this paper. It may be added that their usefulness probably does not extend much further, as they seem to give little help in the problems of space charge and noise. But it appears, from recent investigations of the author, that in many of these problems very good use can be made of Hamilton's third method, which has not been dealt with in this paper. By formulating all equations of the electronic discharge in the form of a single minimum problem, approximate solutions have been found for space-charge problems which otherwise offer great difficulties. It is hoped that it will soon be possible to publish these results.

Appendix I

Proof of Hamilton's Canonical Equations

The first set of the canonical equations constitutes the definition of the total momentum $\mathbf{p}$ according to (4), the second set gives the law of motion. This can be summed up in the vector equation

$$\frac{d\mathbf{p}}{dt} = -\operatorname{grad} \mathcal{H}. \tag{31}$$

Using (5) the left side can be written

$$\frac{d\mathbf{p}}{dt} = \frac{d}{dt}\left(m\mathbf{v} - \frac{e}{c}\mathbf{A}\right)$$
$$= m\frac{\partial \mathbf{v}}{\partial t} - \frac{e}{c}\frac{\partial \mathbf{A}}{\partial t} - \frac{e}{c}(\mathbf{v}\nabla)\mathbf{A} \tag{32}$$

where $(\mathbf{v}\nabla)\mathbf{A}$ is a vector with the c-component

$$v_x\frac{\partial A_x}{\partial x} + v_y\frac{\partial A_x}{\partial y} + v_z\frac{\partial A_x}{\partial z}.$$

By (3) the x-component of the right side of (31) is

$$-\operatorname{grad}_x \mathcal{H} = -\frac{1}{m}\left(p_x + \frac{e}{c}A_x\right)\frac{e}{c}\frac{\partial A_x}{\partial x} + \cdots$$
$$+e\frac{\partial \phi}{\partial x} = \frac{-e}{mc}\left(v_x\frac{\partial A_x}{\partial x} + v_y\frac{\partial A_y}{\partial x} + v_z\frac{\partial A_z}{\partial x}\right) + e\frac{\partial \phi}{\partial x}.$$

By a well-known vector relation, the first term at the right side is the x-component of the vector

$$-\frac{e}{c}\left[(\mathbf{v}\nabla)\mathbf{A} + \mathbf{v}\times \operatorname{curl} \mathbf{A}\right]$$

so that the right side of (31) becomes

$$e \operatorname{grad} \phi - \frac{e}{c}\mathbf{v}\times \operatorname{curl} \mathbf{A} - \frac{e}{c}(\mathbf{v}\nabla)\mathbf{A}. \tag{33}$$

Comparing this with (32) the terms $-(e/c)(\mathbf{v}\nabla)\mathbf{A}$ cancel out. Substituting

$$\mathbf{H} = \operatorname{curl} \mathbf{A} \qquad \mathbf{E} = -\operatorname{grad} \phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}$$

we obtain the law of motion (1) in Lorentz's formulation.

Appendix II

Proof of Lagrange's Theorem

Consider a closed curve c at some instant t, with the circulation

$$C = \oint (p_x\delta x + p_y\delta y + p_z\delta z).$$

The symbol δ will be used in the following for denoting simultaneous distances between particles, while the symbol d will be reserved for their motion. The rate of change of the circulation C during the motion of the curve c is

$$\frac{dC}{dt} = \frac{d}{dt}\oint (p_x\delta x + p_y\delta y + p_z\delta z)$$
$$= \oint\left(\frac{dp_x}{dt}\delta x + \frac{dp_y}{dt}\delta y + \frac{dp_z}{dt}\delta z + p_x\frac{d\delta x}{dt} + p_y\frac{d\delta y}{dt} + p_z\frac{d\delta z}{dt}\right)$$
$$= \oint\left(\frac{dp_x}{dt}\delta x + \frac{dp_y}{dt}\delta y + \frac{dp_z}{dt}\delta z + p_x\delta\frac{dx}{dt} + p_y\delta\frac{dy}{dt} + p_z\delta\frac{dz}{dt}\right)$$
$$= \oint\left[\frac{dp_x}{dt}\delta x + \frac{dp_y}{dt}\delta y + \frac{dp_z}{dt}\delta z + \delta(p_xv_x + p_yv_y + p_zv_z) - (v_x\delta p_x + v_y\delta p_y + v_z\delta p_z)\right]$$
$$= -\oint\left(\frac{\partial \mathcal{H}}{\partial x}\delta x + \frac{\partial \mathcal{H}}{\partial y}\delta y + \frac{\partial \mathcal{H}}{\partial z}\delta z + \frac{\partial \mathcal{H}}{\partial p_x}\delta p_x + \frac{\partial \mathcal{H}}{\partial p_y}\delta p_y + \frac{\partial \mathcal{H}}{\partial p_z}\delta p_z\right) = -\oint \delta\mathcal{H} = 0.$$

In the second line, we have carried out the differentiation. In the third line, the order of displacements along simultaneous positions and along the trajectories has been interchanged. In the fourth, the second term has been transformed so as to split off a complete differential. In the last line, the canonical equations (2) have been used, which transform the whole integrand into a complete differential, the integral of which must vanish along every closed circuit. Thus the circulation is an invariant of the motion.

Appendix III

The Curl of the Total Momentum in Regular Electron Beams

In a regular electron beam, that is to say, in a beam which starts from a cathode with zero velocity, the total momentum $\mathbf{p}$ has an initial vorticity normal to the cathode of the value

$$\operatorname{curl}_n p = -\frac{e}{c}H_n \tag{34}$$

while the tangential components of the curl are zero.

The first part of this statement follows immediately from Lagrange's theorem, if it is applied to any closed circuit of electrons starting simultaneously from the cathode, as at the cathode surface the tangential velocity is zero. In order to prove the second part, assume that the normal n to the cathode surface coincides with the x direction, and the tangential component of H with the z direction. The y component of Lorentz's (1) is

$$m\frac{d}{dt}v_y = m\left(\frac{\partial v_y}{\partial t} + v_x\frac{\partial v_y}{\partial x} + v_y\frac{\partial v_y}{\partial y} + v_z\frac{\partial v_y}{\partial z}\right)$$

$$= -eE_y + \frac{e}{c}v_xH_z. \qquad (35)$$

At the cathode surface

$$v_y = v_z = \frac{\partial v_y}{\partial t} = E_y = 0. \qquad (36)$$

These terms go to zero at the cathode surface at least like x, whereas the remaining terms which have v_x as a factor go to zero like $x^{1/2}$ at a saturated cathode and like $x^{2/3}$ if the discharge is space-charge-limited. Thus the remaining terms are dominant near the cathode surface. Dividing these by v_x we obtain

$$m\frac{\partial v_y}{\partial x} = \frac{e}{c}H_z. \qquad (37)$$

On the other hand at the cathode surface, as $v_x = 0$

$$\text{curl}_z\,\mathbf{v} = \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} = \frac{\partial v_y}{\partial x}.$$

Thus

$$\text{curl}_z\,\mathbf{p} = \text{curl}_z\left(m\mathbf{v} - \frac{e}{c}\mathbf{A}\right)$$

$$= \text{curl}_z\,m\mathbf{v} - \frac{e}{c}H_z = 0 \qquad (38)$$

that is to say, the tangential component of the vorticity of the total momentum at the surface of a cathode is zero in all circumstances.

The Plane-Wave Resolution of Guided Waves*

S. S. MACKEOWN†, FELLOW, I.R.E., AND JOHN W. MILES‡

Summary—**Wave propagation in cylindrical guides of both rectangular and circular cross section is treated by representing the proper solutions to Maxwell's equation through a plane-wave expansion of the Hertzian vector. In the case of a rectangular wave guide only a finite number of plane waves (two or four) is required to represent a given mode, while for the circular guide an infinite manifold is required. The plane waves are uniform, traveling with the velocity of light in the medium at an angle to the cylindrical axis which is determined by the frequency and the eigenvalue of the mode under consideration.**

Introduction

THE PROPOSITION that any solution to the wave equation may be resolved into an expansion of plane waves has been enunciated by several writers, including Stratton[1] and Ramo and Whinnery.[2] For academic purposes it nevertheless seems worth while to produce the expansions explicitly for the important cases of propagation in circular and rectangular wave guides. These expansions and concepts have proved helpful in classroom presentation of the subject, particularly in clarifying the concepts of phase and group velocities.

The Field Equations

In the interests of simplicity we shall derive all field vectors from the Hertzian vector, following Stratton.[1] In cylindrical co-ordinates (u, v, z), the fields are completely specified by the z component of the Hertzian vector which is herein represented by the scalar function $\psi(u, v, z)$. (z is the co-ordinate measured along the axis of the cylinder, and u and v are orthogonal co-ordinates in a plane of constant z.) Inasmuch as the Hertzian vector satisfies the vector wave equation and the z co-ordinate is measured in a fixed direction, its z component must satisfy the scalar wave equation; viz.,

$$\nabla^2\psi + k^2\psi = 0, \quad k = (\mu\epsilon)^{1/2}\omega = 2\pi/\lambda \qquad (1)$$

where μ and ϵ are the permeability and the dielectric constant, respectively, ω is the angular frequency, and λ is the wave length in the medium (not, however, the guide wave length). The operator $\partial/\partial t$ has been replaced by $j\omega$ in accordance with the assumption of harmonic time variation.[3]

The scalar and vector potentials ϕ and $\overline{A}$ are then given by[4]

$$\phi = -\partial\psi/\partial z \qquad (2)$$

$$\overline{A} = j\omega\mu\epsilon\bar{k}\psi. \qquad (3)$$

In the case of TM or E waves; i.e., $H_z = 0$, we have

$$\overline{E}_E = -\nabla\phi_E - j\omega\overline{A}_E \qquad (4)$$

$$\overline{B}_E = \nabla \times \overline{A}_E \qquad (5)$$

while for TE or H waves ($E_z = 0$) we have

$$\overline{D}_H = -\nabla \times \overline{A}_H \qquad (6)$$

$$\overline{H}_H = -\nabla\phi_H - j\omega\overline{A}_H. \qquad (7)$$

The boundary conditions require the vanishing of the

* Decimal classification: R116. Original manuscript received by the Institute, February 6, 1945; revised manuscript received, April 16, 1945.

† California Institute of Technology, Pasadena, Calif.
‡ Lockheed Aircraft Corporation, Burbank, California.

[1] J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., New York, N. Y., 1941.

[2] Simon Ramo and John Whinnery, "Fields and Waves in Modern Radio," John Wiley and Sons, New York, N. Y., 1944.

[3] Note that Stratton assumes a time variation $e^{-i\omega t}$ where we have assumed $e^{j\omega t}$. Simply substituting $-j$ for i reconciles the two assumptions, $e^{j\omega t}$ being more general engineering usage.

[4] See Appendix A for a resume of the vector notation used herein.

tangential electric field or, equivalently, of the normal magnetic field at the metal surface. For the E modes we may require $(\nabla \times \overline{A}_E) \cdot \bar{n}$ to vanish,[5] which, from (3) is effected by the vanishing of the tangential derivative of ψ_E; viz.,

$$\partial \psi_E / \partial s = 0 \tag{8}$$

while for the H modes we require $(\nabla \times \overline{A}_H) \cdot \bar{s}$ to vanish, which from (3) is effected by the vanishing of the normal derivative of ψ_H; viz.,

$$\partial \psi_H / \partial n = 0. \tag{9}$$

In treating (1) it is customary to assume propagation along the z axis with a propagation constant jh so that we may write

$$\psi(u, v, z) = f(u, v) e^{\mp jhz} \tag{10}$$

where the positive or negative sign is associated with propagation in the positive or negative z direction, respectively. We remark that the mathematical motivation of this choice is one of simplicity, and is not necessarily accompanied by any a priori conviction as to the direction of travel of any individual wave front. The wave represented in (10) has a phase velocity (ω/h). Substituting (10) in (1) we have the two-dimensional wave equation

$$\nabla^2 f + \kappa^2 f = 0, \qquad \kappa^2 = k^2 - h^2. \tag{11}$$

There is a doubly infinite set of solutions to (11) corresponding to a doubly infinite set of discrete values for κ, these eigenvalues being determined by the boundary conditions (8) or (9).

The Rectangular Wave Guide

The solutions to (1), satisfying the boundary conditions (8) and (9) in a rectangular wave guide bounded by the planes $x=0$, $x=a$, $y=0$, and $y=b (u=x, v=y)$, are found to be (see Appendix B)

$$\psi_{E_{mn}}(x, y, z) = \sin(m\pi x/a) \sin(n\pi y/b) e^{\mp jhz} \tag{12}$$

$$\psi_{H_{mn}}(x, y, z) = \cos(m\pi x/a) \cos(n\pi y/b) e^{\mp jhz} \tag{13}$$

$$\kappa_E{}^2 = \kappa_H{}^2 = (m\pi/a)^2 + (n\pi/b)^2 \tag{14}$$

where m and n are integers. We observe that, by elementary trigonometry, both (12) and (13) may be obtained by superimposing solutions to (1) of the type

$$\psi_{mn}(x, y, z) = e^{j[\pm m\pi(x/a) \pm n\pi(y/b) \pm hz]} \tag{15}$$

where any combination of the signs is permitted. If we define the angles

$$\alpha = \tan^{-1}(\kappa/h) = \sin^{-1}(\kappa/k) \tag{16}$$

$$\beta = \pm \tan^{-1}(mb/na) \tag{17}$$

we may write

$$\psi_{mn}(x, y, z) = e^{-j\overline{K} \cdot \overline{R}} \tag{18}$$

$$\overline{K} = [\bar{\imath} \sin\alpha \cos\beta + \bar{\jmath} \sin\alpha \sin\beta + \bar{k} \cos\alpha] k \tag{19}$$

$$\overline{R} = \bar{\imath} x + \bar{\jmath} y + \bar{k} z \tag{20}$$

$\overline{K}$ being the "vector-propagation constant" and $\overline{R}$ the radius vector. The function given by (18) represents a uniform plane wave traveling at a polar angle α to the z axis and an azimuthal angle β to the X axis with the velocity of light $(c = (\mu\epsilon)^{-1/2})$ in the medium. The properties of a plane wave, listed in any standard test,[1,2] include the mutual orthogonality of the electric field, the magnetic field, and the axis of propagation and, for a uniform plane wave, the uniformity of the fields in any plane transverse to the axis of propagation. From (16) we observe that the phase velocity (ω/h) of any mode is greater than the velocity of light by a factor of $\sec\alpha$. Accordingly, the measured guide wavelength is $\sec\alpha$ greater than λ, inasmuch as it is measured between planes of constant phase; however, since the plane waves which make up the guided wave travel at an angle α to the z axis, the group velocity of the guided wave is less than the velocity of light by a factor of $\cos\alpha$.

Chu and Barrow,[6] following Brillouin and Page and Adams, have given the plane-wave resolution for the special case of the TE_{01} mode, where b is greater than a, and mention the experimentally observed phenomenon that radiation through a slot cut in one face of the guide is at the predicted angle $\alpha = \sin^{-1}(\lambda/2b)$ (see (14) and (16) for $m=0$, $n=1$) to the z axis. As a more explicit example of a plane-wave resolution in a rectangular guide, we choose the H_{mn} mode traveling along the positive z axis; from (13) we may write

$$\begin{aligned}\psi_{H_{mn}}(x, y, z) = 1/4\big[&e^{-j(m\pi x/a + n\pi y/b + hz)} \\ &+ e^{-j(-m\pi x/a + n\pi y/b + hz)} \\ &+ e^{-j(-m\pi x/a - n\pi y/b + hz)} \\ &+ e^{-j(m\pi x/a - n\pi y/b + hz)}\big] \\ = 1/4[&\psi_1 + \psi_2 + \psi_3 + \psi_4]\end{aligned} \tag{21}$$

where ψ_1, ψ_2, ψ_3, and ψ_4 are plane waves traveling at polar and azimthal angles

$$\alpha = \sin^{-1}\left\{[(m/a)^2 + (n/b)^2]^{1/2} \lambda/2\right\}$$

$$\beta_4^1 = \pm \tan^{-1}(mb/na), \qquad \beta_3^2 = \pi \mp \beta_1 \tag{22}$$

respectively. We remark that, for the special case $m=0$ $\psi_1 = \psi_4 (\beta_1 = \beta_4 = 0)$ and $\psi_2 = \psi_3 (\beta_2 = \beta_3 = \pi)$, while for $n=0$ $\psi_1 = \psi_2 (\beta_1 = \beta_2 = \pi/2)$ and $\psi_3 = \psi_4 (\beta_3 = \beta_4 = 3\pi/2)$.

Circular Guide

For the case of the circular guide we use a treatment which is suggested by Stratton's general treatment of cylindrical waves.[1] The solutions to (1) in the co-ordinates $(u=r, v=\theta, z)$ satisfying the boundary conditions (8) and (9) in a circular guide of radius a are found to be (see Appendix B)

$$\psi_{mn}(r, \theta, z) = J_m(\kappa_n r) e^{j(\pm m\theta + \phi \mp hz)} = f(r, \theta) e^{\mp jhz} \tag{23}$$

$$J_m(\kappa_n{}^E a) = 0 \tag{24}$$

$$J_m{}'(\kappa_n{}^H a) = 0 \tag{25}$$

where $Jm(x)$ is the Bessel function of the m'th order.[7] In general, several solutions of the type (23) will have to be linearly combined to give the correct polarization in θ. By expressing the Bessel function as the contour integral[8] and making the change of variable $\beta = \phi \pm \theta$

[5] $\bar{n}$ is the unit vector normal to the cylindrical surface in question; similarly $\bar{s}$ is the unit tangential vector in a plane of constant z.

[6] L. J. Chu and W. L. Barrow, "Electromagnetic waves in hollow metal tubes of rectangular cross section," Proc. I.R.E., vol. 26, pp. 1520–1555; December, 1938.

[7] ϕ is an arbitrary phase constant specifying the polarization of the potential.

[8] Eugene Jahnke and Fritz Emde, "Tables of Functions," Dover Publications, New York, N. Y., 1943, p. 147.

$$J_m(\kappa_n r) = (j^{-m}/2\pi)\int_0^{2\pi} e^{-j[\kappa_n r \cos\phi + m\phi]}d\phi$$
$$= \int_{\pm\theta}^{2\pi\pm\theta} e^{-j[\kappa_n r \cos(\beta\mp\theta)+m(\beta\mp\theta)-m\pi/2]}d\beta. \quad (26)$$

Hence

$$f(r,\theta) = \int_{\pm\theta}^{2\pi\pm\theta} e^{-jk\sin\alpha\cos(\beta\mp\theta)r}d\beta \quad (27)$$
$$g(\beta) = 1/2\pi e^{-jm(\beta+\pi/2)} \quad (28)$$

where α is given by (16). Expanding the cosine in the exponent of (27) and substituting in (23) we obtain

$$\psi(r,\theta,z) = \int_{\pm\theta}^{2\pi\pm\theta} g(\beta)e^{-j\overline{K}\cdot\overline{R}}d\beta \quad (29)$$

where $\overline{K}$ and $\overline{R}$ are given by (19) and (20).

The solution given by (29) represents an infinite manifold of plane waves, having amplitudes $d\beta/2\pi$ and phase $2\pi g(\beta)$ propagating along coaxial cones of aperture 2α and at azimuthal angles β with the velocity of light in the medium. The situation is more complex than in the case of the rectangular guide since the number of waves required for the resolution of a guided wave is infinite, but since the angle α is given by (16) in both cases, the phase velocity, the group velocity, and the phenomenon of cutoff can all be interpreted as in the case of the rectangular guide.

As an example of plane-wave resolution in a circular guide, we choose a vertically polarized H_{11} wave, since this mode has the lowest cutoff frequency of all possible circular modes, traveling in the positive z direction. From (3) and (6) we conclude that vertical polarization; i.e., no θ component of electric field at $\theta=\pi/2$, requires the solutions of (23) to be combined in such a way to give ψ varying as $\cos\theta$; hence we write

$$\psi_{H_{11}}(r,\theta,z) = 1/2J_1(\kappa_1 r)[e^{j\theta}+e^{-j\theta}]e^{-jhz}$$
$$= 1/2(\psi_1+\psi_2) \quad (30)$$
$$J_1'(\kappa_1 a) = 0 \quad (31)$$

where $$\psi_2^1 = \int_{\pm 0}^{2\pi\pm\theta}\left[\frac{je^{-j\beta}}{2\pi}\right]e^{-jk\sin\alpha\cos(\beta\mp\theta)r}d\beta. \quad (32)$$

ψ_2^1 represent two infinite manifolds of plane waves of phase distribution $je^{-j\beta}$ and amplitudes $1/2\pi d\beta$, each manifold being constituted to effect the vanishing of the tangential electric field at $r=a$, and the two manifolds being combined to give a vertically polarized wave.

A similar treatment may be made for the solutions in a coaxial guide if the contour integral representation of the Neumann function is introduced along with the representation of the Bessel function given by (26). When the Bessel and Neumann functions are combined in such a way that, for the proper choice of the eigenvalue, the tangential electric field vanishes at both the inner and outer conductors, the resultant manifold of plane waves will be found to be reflected back and forth, while also progressing along the z axis, between the outer and inner conductors.

Appendix A

Vector Notation

The co-ordinate systems used in the foregoing were rectangular (right-handed) and cylindrical polar (θ measured counterclockwise and $\theta=0$ when $\bar{r}_1\equiv\bar{\imath}$), and the corresponding vector co-ordinates are ($\bar{x}=\bar{\imath}x$, $\bar{y}=\bar{\jmath}y$, $\bar{z}=\bar{k}z$) and ($\bar{r}=\bar{r}_1 r$, $\bar{\theta}=\bar{\theta}_1\theta$, $\bar{z}=\bar{k}z$) where $\bar{\imath}$, $\bar{\jmath}$, $\bar{k}$, $\bar{r}_1$, and $\bar{\theta}_1$ are unit vectors in the positive x, y, z, r, and θ directions, respectively.

The gradient, or directional derivative of a scalar V is given by

$$\nabla V = \bar{\imath}\partial V/\partial x + \bar{\jmath}\partial V/\partial y + \bar{k}\partial V/\partial z \quad (33)$$

which, in cylindrical polar co-ordinates, may be written

$$\nabla V = \bar{r}_1\partial V/\partial r + \bar{\theta}_1/r\,\partial V/\partial\theta + \bar{k}\partial V/\partial z. \quad (34)$$

The curl of a vector $\overline{V}(=\bar{\imath}V_x+\bar{\jmath}V_y+\bar{k}V_z)$ is given by

$$\nabla\times\overline{V} = \bar{\imath}(\partial V_z/\partial y - \partial V_y/\partial z) + \bar{\jmath}(\partial V_x/\partial z - \partial V_z/\partial x)$$
$$+ \bar{k}(\partial V_y/\partial x - \partial V_x/\partial y) \quad (35)$$

which, in cylindrical polar co-ordinates, may be written

$$\nabla\times\overline{V} = \bar{r}_1\left[\frac{1}{r}\frac{\partial V_z}{\partial\theta} - \frac{\partial V_\theta}{\partial z}\right] + \bar{\theta}_1\left[\frac{\partial V_r}{\partial z} - \frac{\partial V_z}{\partial r}\right]$$
$$+ \bar{k}\left[\frac{1}{r}\frac{\partial(rV_\theta)}{\partial r} - \frac{\partial V_r}{\partial\theta}\right]. \quad (36)$$

Appendix B

Solutions of the Scalar-Wave Equation

The scalar-wave equation[8] may be written

$$\partial^2\psi/\partial x^2 + \partial^2\psi/\partial y^2 + \partial^2\psi/\partial z^2 + k^2\psi = 0. \quad (37)$$

The assumption of a cylindrical-co-ordinate system (u, v, z) can be introduced by assuming a solution of the form (10), so that the solution of (37) is reduced to the solution of (11), which may be written

$$\partial^2 f/\partial x^2 + \partial^2 f/\partial y^2 + \kappa^2 f = 0. \quad (38)$$

On rectangular co-ordinates the substitution of the trial solution $f(x,y)=X(x)Y(y)$ leads to the result

$$f(x,y) = (A\cos\mu x + B\sin\mu x)(C\cos\nu y + D\sin\nu y) \quad (39)$$
$$\kappa^2 = \mu^2+\nu^2. \quad (40)$$

If the planes $x=0$, $x=a$, $y=0$, and $y=b$ are assumed perfectly conducting the boundary condition (8) requires

$$\left.\frac{\partial\psi_E}{\partial x}\right|_{y=0} = \left.\frac{\partial\psi_E}{\partial x}\right|_{y=b} = \left.\frac{\partial\psi_E}{\partial y}\right|_{x=0} = \left.\frac{\partial\psi_E}{\partial y}\right|_{x=a} = 0 \quad (41)$$

while (9) requires

$$\left.\frac{\partial\psi_H}{\partial x}\right|_{x=0} = \left.\frac{\partial\psi_H}{\partial x}\right|_{x=a} = \left.\frac{\partial\psi_H}{\partial y}\right|_{y=0} = \left.\frac{\partial\psi_H}{\partial y}\right|_{y=b} = 0. \quad (42)$$

From (10) it is seen that (41) and (42) may be applied directly to $f(u,v)$, in this case given by (39). Applying (41) to (39) we obtain

$$f_E(x,y) = \sin\mu x\sin\nu y,\ \mu = m\pi/a,\ \nu = n\pi/b \quad (43)$$

while applying (42) to (39) yields

$$f_H(x,y) = \cos\mu x\cos\nu y,\ \mu = m\pi/a,\ \nu = n\pi/b. \quad (44)$$

Substituting (43) and (44) in (10) yields (12), (13), and (14).

In polar co-ordinates (r, θ), (38) becomes

$$\partial^2 f/\partial r^2 + 1/r\, \partial f/\partial r + 1/r^2\, \partial^2 f/\partial\theta^2 + \kappa^2 f = 0. \quad (45)$$

The assumption of the trial solution $f(r, \theta) = R(r)\Theta(\theta)$ yields Bessel's equation for $R(r)$ and the harmonic equation for $\Theta(\theta)$ with the result

$$f(r, \theta) = [AJ_m(\kappa r) + BN_m(\kappa r)]e^{j(\pm m\theta + \phi)} \quad (46)$$

where J_m and N_m are Bessel functions of the first and second kind, respectively, and A, B, and ϕ are arbitrary. Inasmuch as $N_m(0)$ is infinite, B must be equated to zero for a solution in a hollow circular guide (this would not be true in a coaxial guide), while continuity of the solution with respect to θ demands that m be an integer. Assuming the perfectly conducting boundary $r = a$, (8) and (9) become

$$\partial\psi_E/\partial\theta \big|_{r=a} = 0. \quad (47)$$

$$\partial\psi_H/\partial r \big|_{r=a} = 0. \quad (48)$$

Applying these conditions to (46), the transcendental equations (24) and (25) are obtained, where the integer n denotes the sequence of the roots to the equations starting with the smallest in each case.

By convention, n runs from one to infinity, and there is no $n = 0$ root.

Note on the Measurement of Transformer Turns-Ratio*

P. M. HONNELL†, SENIOR MEMBER, I.R.E.

Summary—**This note shows that in many important cases the turns-ratio of iron-core transformers is given by the simple relation**

$$N_1/N_2 = \sqrt{X_{sc1}/X_{sc2}},$$

in which X_{sc1} and X_{sc2} are the reactive components of the short-circuit primary and secondary impedance of the windings concerned. The measurements are conveniently made with an impedance or inductance bridge. This equation also gives the turns-ratio of certain air-core transformers.

I. Introduction

THE RATIO of turns of its windings is probably the most important parameter of a transformer; and the occasion for measuring this ratio for transformers at hand is of frequent occurrence.

The procedure[1–4] most commonly given for determining this parameter is that of exciting the primary winding by a known alternating voltage, and measuring the secondary winding open-circuit voltage by means of a high-impedance voltmeter. The turns-ratio is then taken as equal to the ratio of the terminal voltages of the windings. Although this method is very useful, particularly with power transformers, its accuracy is seriously limited by leakage inductance and distributed capacitance in the windings, and by instrumentation This is particularly true when use is made of vacuum-tube voltmeters (to minimize the secondary burden) in the measurement of audio-frequency transformers. Even if not subject to these errors, this method is only an approximation, since it is strictly a measure of the ratio of the transformer mutual impedance to primary impedance, and not a measurement of the turns-ratio.

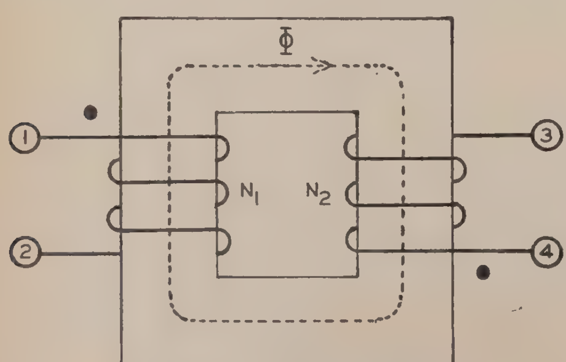


Fig. 1—The iron-core transformer.

Precision methods of calibrating instrument-transformer ratios[5] are likewise concerned only with determinations of *terminal* voltage or current ratios, and are also not measurements of the actual turns-ratio.

The purpose of this paper is to point out the simplicity of determining the turns-ratio by measuring the transformer short-circuit impedances from both the primary winding (giving Z_{sc1}) and the secondary winding (giving Z_{sc2}). These measurements are conveniently made on an impedance or inductance bridge. The desired transformer turns-ratio N_1/N_2 is given to a high order of accuracy by the square root of the quotient of the reactive (or inductive) components of the short-circuit impedances; that is,

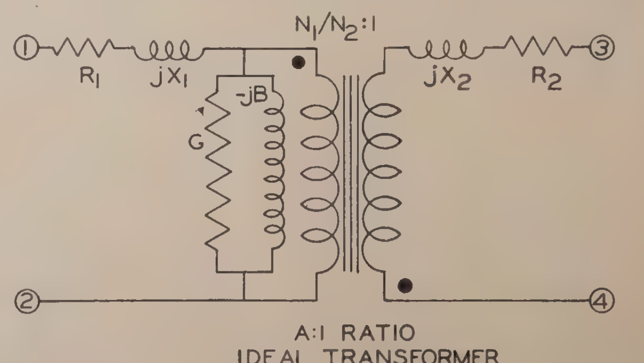


Fig. 2—An equivalent circuit representing the iron-core transformer of Fig. 1.

$$N_1/N_2 = \sqrt{X_{sc1}/X_{sc2}} = \sqrt{L_{sc1}/L_{sc2}}. \quad (1)$$

This method—for all its ease of application and high accuracy—is apparently not well known.

II. The Iron-Core Transformer

The basis for the application of (1) to an iron-core transformer—such as shown in Fig. 1—may be established from a consideration of one of its equivalent circuits, shown in Fig. 2. In the latter illustration,

* Decimal classification: R264.3. Original manuscript received by the Institute, April 17, 1945; revised manuscript received, June 21, 1945.

† Signal Corps, U. S. Military Academy, West Point, New York.

[1] F. E. Terman, "Radio Engineers' Handbook," first edition, McGraw-Hill Book Company, New York 18, N. Y., 1943, p. 973.

[2] A. E. Knowlton, "Standard Handbook for Electrical Engineers," McGraw-Hill Book Company, New York 18, N. Y., 1941, p. 611.

[3] H. Pender and K. McIlwain, "Electrical Engineers' Handbook," John Wiley and Sons, Inc., New York, N. Y., 1941, pp. 7–23, section 7.

[4] Electrical Engineering Staff, Massachusetts Institute of Technology, "Magnetic Circuits and Transformers," John Wiley and Sons, Inc., New York, N. Y., 1943, pp. 442–448, 352–353.

[5] F. A. Laws, "Electrical Measurements," Second Edition, McGraw-Hill Book Company, Inc., New York, 1938, pp. 609–622.

N_1 and N_2 = the number of primary and secondary turns,
R_1 and R_2 = the effective resistances of the windings,
X_1 and X_2 = the *leakage reactances* of the windings,
G and B = the equivalent conductance and susceptance due to the magnetic characteristics of the iron core referred to the primary winding, and
a = the ratio of transformation or turns-ratio N_1/N_2 of the transformer windings under consideration.

The short-circuit impedance of the transformer as measured between the primary terminals 1 and 2, with the secondary terminals 3 and 4 short-circuited, is

$$Z_{sc1} \equiv R_{sc1} + jX_{sc1} = R_1 + jX_1 + \frac{a^2(R_2 + jX_2)}{1 + a^2(R_2 + jX_2)(G - jB)}. \quad (2)$$

Likewise, the secondary short-circuit impedance measured between terminals 3 and 4, with the primary terminals 1 and 2 short-circuited, is

$$Z_{sc2} \equiv R_{sc2} + jX_{sc2} = R_2 + jX_2 + \frac{1}{a^2}\cdot\frac{R_1 + jX_1}{1 + (R_1 + jX_1)(G - jB)}. \quad (3)$$

Equations (2) and (3) yield the following reactive components, respectively:

$$X_{sc1} = X_1 + a^2X_2 \cdot \frac{[1+a^2(GR_2+BX_2)] - (R_2/X_2)[a^2(GX_2-BR_2)]}{[1+a^2(GR_2+BX_2)]^2 + [a^2(GX_2-BR_2)]^2} \quad (4)$$

$$X_{sc2} = X_2 + \frac{X_1}{a^2}\cdot\frac{[1+GR_1+BX_1] - (R_1/X_1)[GX_1-BR_1]}{[1+GR_1+BX_1]^2 + [GX_1-BR_1]^2}. \quad (5)$$

Case 1. Unrestricted: Under the most general conditions, the ratio of (4) to (5) yields a symmetrical but unwieldy expression which apparently does not yield the results indicated by the simple equation (1). Recourse to experiment, however, discloses that (1) *does* correctly give the turns-ratio of transformers to the limit of the accuracy obtainable with impedance bridges. The reasons for this are perhaps more evident when the following special cases are considered.

Case 2. $G=0$, $B=0$: With $G=0$ and $B=0$, the ratio of (4) to (5) gives

$$X_{sc1}/X_{sc2} = [(X_1 + a^2X_2)/(X_2 + (X_1/a^2))] = a^2.$$

Thus,[6] we have finally the exact equation

$$N_1/N_2 \equiv a = \sqrt{X_{sc1}/X_{sc2}} = \sqrt{L_{sc1}/L_{sc2}}. \quad (1)$$

Now the condition $G=0$ and $B=0$ is tantamount to stipulating that the transformer is free of core loss and that it requires no magnetizing current. Since X_{sc1} and X_{sc2} are *short-circuit* parameters, this condition is very closely approximated because the core may be operated at very low flux densities during the measurements. As easily demonstrated experimentally, the transformer then acts like a perfectly linear device, and the usual large variations of inductance with current amplitude and with frequency are greatly reduced and of the same order of magnitude as the variations in an air-core inductor.

Case 3. $(G-jB)\neq 0$, *but* $a^2R_2=R_1$ *and* $a^2X_2=X_1$: Under these conditions core losses and magnetizing current may exist, but (4) and (5), or more simply (2) and (3), also yield the *exact* relation[6]

$$N_1/N_2 \equiv a = \sqrt{X_{sc1}/X_{sc2}} = \sqrt{L_{sc1}/L_{sc2}}. \quad (1)$$

Now the restrictions imposed here are those commonly satisfied in transformers used for power transfer: that is, "impedance matching" or "power" transformers. In such transformers, the winding volume is divided equally and uniformly between the several windings in order to minimize copper losses and leakage inductances. This automatically insures[7] that $a^2R_2=R_1$ and $a^2X_2=X_1$.

Thus it is evident that, in important practical cases, the turns-ratio of transformers may be determined from the measured values of short-circuit reactances (or inductances) and substitution in (1). Furthermore, the effects of distributed capacitance in the windings is minimized due to the reduced impedance level of the measurements. As for the winding leakage reactances, instead of being a source of error, they make possible this very convenient and accurate method of measuring transformer turns-ratio.

III. Air-Core Transformers

If we define an *air-core transformer* as that special class of inductively coupled circuit which has identical winding dimensions, volume, and shape for the primary and secondary windings, then experiment will show that the expression

$$N_1/N_2 = \sqrt{X_{sc1}/X_{sc2}} \quad (1)$$

also furnishes a good approximation to the turns-ratio. This may be shown from the following development.

The reactive component X_{sc1} of the short-circuit impedance measured at the primary terminals 1 and 2 of the inductively coupled coils shown in Fig. 3, with terminals 3 and 4 short-circuited, is

$$X_{sc1} = \omega L_1 - [(\omega^2M^2)/(R_2^2 + (\omega L_2)^2)]\cdot\omega L_2$$

in which the symbols have their usual meanings. This may be written

$$X_{sc1} = \omega L_1[1 - k^2/(1 + (1/Q_2)^2)] \quad (6)$$

where $Q_2 = \omega L_2/R_2$ is the reactance-resistance ratio of the secondary coil alone, and $k^2 = M^2/L_1L_2$ is the coefficient of coupling.

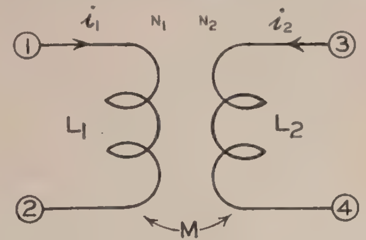


Fig. 3—The air-core transformer.

Similarly, the reactive component X_{sc2} of the short-circuit impedance measured at the secondary terminals 3 and 4 with terminals 1 and 2 short-circuited, is

$$X_{sc2} = \omega L_2[1 - k^2/(1 + (1/Q_1)^2)] \quad (7)$$

where $Q_1 = \omega L_1/R_1$ is the reactance-resistance ratio of the primary coil alone. The ratio of (6) to (7) gives

$$X_{sc1}/X_{sc2} = L_1/L_2\cdot[(1-k^2/(1+(1/Q_2)^2))/(1-k^2/(1+(1/Q_1)^2))]. \quad (8)$$

Now the turns-ratio of coupled coils is related to their inductance by the relation[8]

$$L_1/L_2 = (N_1/N_2)^2\cdot k_2/k_1. \quad (9)$$

In this equation, the primary and secondary coupling factors k_1 and k_2 are defined as follows:

$$k_1 \equiv \phi_{21}/\phi_{11} \quad (10)$$
$$k_2 \equiv \phi_{12}/\phi_{22}. \quad (11)$$

in which

ϕ_{11} = the equivalent total flux linking turns N_1 due to the current i_1 acting alone,
ϕ_{22} = the equivalent total flux linking turns N_2 due to the current i_2 acting alone,
ϕ_{21} = the equivalent mutual flux linking N_2 due to the current i_1 in N_1 acting alone,
ϕ_{12} = the equivalent mutual flux linking N_1 due to the current i_2 in N_2 acting alone.

From the definitions (10) and (11), and the usual definitions of self and mutual inductance $L_1 \equiv N_1\phi_{11}/i_1$, $L_2 \equiv N_2\phi_{22}/i_2$, and $M \equiv N_2\phi_{21}/i_1 = N_1\phi_{12}/i_2$, equation (9) is easily derived. (It can also easily be shown that the coefficient of coupling is related to the coupling factors by the expression $k^2 = k_1\cdot k_2$.)

Substituting (9) into (8) and solving for the turns-ratio gives

$$\frac{N_1}{N_2} = \sqrt{\frac{X_{sc1}}{X_{sc2}}}\sqrt{\frac{k_1}{k_2}}\sqrt{\frac{1 - k^2/(1 + (1/Q_1)^2)}{1 - k^2/(1 + (1/Q_2)^2)}} \quad (12)$$

which applies to *all* coupled coils. The direct application of (12) to the determination of the turns-ratio of coupled coils is unfortunately not possible since k_1 and k_2 cannot be determined from measurements, although experimental confirmation of the equation is simple enough when the turns-ratio is known.

For *air-core transformers* as defined above, the primary and secondary volumes are indistinguishable except as to the number of turns (and if the winding space factors are identical and the conductors finely divided so that eddy-current losses are unimportant), then[9] $Q_1 = Q_2$, and also by symmetry, $k_1 = k_2$. As a result, (12) reduces to

$$N_1/N_2 = \sqrt{X_{sc1}/X_{sc2}} \quad (1)$$

and is *exact* for air-core transformers as herein defined. It should be noted that (1) also applies in this case even though the magnetic coupling between the windings of the transformer approaches zero; that is, the coils are not coupled together at all!

[6] Note that under these conditions the turns-ratio is also given by $N_1/N_2 = \sqrt{R_{sc1}/R_{sc2}}$.
[7] See page 342 of footnote reference 4.

[8] This equation is derived in Electrical Engineering Staff, Massachusetts Institute of Technology, "Magnetic Circuits and Transformers," John Wiley and Sons, Inc., New York, N. Y., 1943, chapter 17.
[9] See pages 216–218 of footnote reference 4.

Discussion on

"The Theory of Transmission Lines"*

EDWARD N. DINGLEY, JR.

Fred J. Heath:[1] Mr. Dingley has made a very useful contribution towards the understanding of transmission-line theory by those who are not familiar with the mathematics so often used. In particular, the use of the exponential notation throughout, and the use of Fig. 2, with its accompanying text, makes it easy to visualize the formation of standing waves, and the occurrence of other phenomena on transmission lines.

Because of the fact that this is a "tutorial" paper, it seems desirable to point out those parts which appear to be misleading to the student. At the risk of appearing pedantic, I am listing such parts in the order in which they appear in Mr. Dingley's paper, with my comments on each.

Page 118, Paragraph 1:

The words "infinitely long" constitute an unnecessary restriction of the first statement. It would be desirable to confine reference to the length of the line to those cases in which it is of particular significance.

Page 119, Paragraph 2, discussion of (5) and (5a):

Might it not have been better to say, "If (5) and (5a) are true, they must be true at all points along the line, including the point where $l=0$." It seems undesirable to talk of a line of zero length, when a line of unspecified, but fixed, length is easier to visualize, and is satisfactory from a mathematical standpoint.

The subscript s in (6) and (6a) refers to the sending end of the line, *where* $l=0$.

Page 119, Paragraph 4, (8) and (8a):

Again, it might have been better to say, "In (7) and (7a), at the point $l=0$,

$$dE/(dl) = \cdots \quad \text{(at } l = 0) \tag{8}$$

and

$$dI/(dl) = \cdots \quad \text{(at } l = 0). \tag{8a}$$"

Mr. Dingley's meaning is clear, but E_s and I_s are not functions of l in this particular case, where the constants of integration in (5) and (5a) are being evaluated. It may, in fact, be considered that (5) to (13a), inclusive, are being considered as related to a line of unspecified, but fixed, length, with a fixed load at the far end, and a fixed generator supplying E_s and I_s at the point $l=0$.

* Proc. I.R.E., vol. 33, pp. 118–125; February, 1945.
[1] Research Enterprises, Ltd., Leaside, Ont., Canada.

Once (13) and (13a) have been obtained, they may be applied to *any* line.

Page 119, Paragraph 10, (14) and (14a):

The true significance of "l" is not brought out in the development of these equations. It is suggested that it might have been better to say, "If the line is infinitely long, then E and I at the far end (where $l=\infty$) must be zero."

Page 119, Paragraph 11, (16) and (16a):

The reference to an "infinitely long" line at this point is particularly unfortunate, following as it does the derivation of (15) and (15a). In view of the fact that the main development from here on is based on (16) and (16a), it is unfortunate that Mr. Dingley did not deduce these equations directly from the basic equations, rather than through a loose transformation of (13) and (13a). The only apparent use made of (13) and (13a) is the deduction of the value of Z_0. This could have been obtained with equal facility by dividing (16) by (16a) and obtaining the limit, as $l\rightarrow\infty$, of E/I.

Page 120, Paragraph 8, discussion of α and β:

The nature of β, and its relation to an angle, measured in circular radians, is clearly shown in the latter part of the paper. The relation of α to the "hyperbolic radian" is passed over with very little comment. One recognizes, of course, the necessity of limiting the scope of such a paper as this, yet, if the hyperbolic angle is to be brought into the discussion at all; it would seem best to show the connection between the exponential form of the equations, used throughout the paper, and the hyperbolic form, which is used so frequently in the standard texts. Perhaps a footnote reference to some of these texts would be sufficient.

Page 120, Paragraphs 8 and 13:

The statement that "α represents the *change in amplitude* of the voltage or current as it travels a unit length of transmission line" is presumably based on an infinite line, or one with a load equal to Z_0, as it neglects the effect of standing waves. Strictly speaking, α represents the change in the *logarithm* of the amplitude of the voltage or current as it travels a unit length of the transmission line. As such, it is a measure of the *ratio* of the amplitude of the voltage (or current) after traveling the unit length of the transmission line to its amplitude on entering the unit length of line.

Page 121, Paragraphs 3 and 5, following (40) and (46):

In both these paragraphs, it should be emphasized that not only is $R=0=G$ but also $K=1=\mu$.

Page 121, Paragraph 5, following (46):

The statement "The sending and receiving voltages (and currents) are equal" is misleading. The whole of the *input power* is received at the load, without loss in the line, but the actual values of voltage and current at the two ends of the line may differ considerably, due to the standing waves mentioned at the end of the paragraph.

Page 121, Paragraph 6:

Not only must $R=0=G$, but also the product $K\times\mu$ must be independent of frequency.

Page 121, Paragraph 8, following (48):

While it is true that (47) is the same as (32), Mr. Dingley has overlooked the fact that (45) is obtained through (41) from (37) and (39), which are based on the condition that $K=1=\mu$.

As stated by Mr. Dingley, the condition for (47), that $R\times B=G\times X$, is generally obtained by loading with either lumped inductors, or by uniform wrapping of the line with high-permeability material. The velocity of transmission under these conditions may be deduced as follows:

Substitute from (47) in (30)

$$V=\omega/\beta=\omega/\omega\sqrt{LC}=1/\sqrt{LC} \text{ unit lengths per second.} \quad (1)$$

The loading of the cable increases the effective inductance of the cable. This may be accounted for by setting μ in (35) equal to $\mu_1>1$. In practical cables, K in (36) is >1, and will here be set equal to $K_1>1$.

If (1) is applied to a practical, loaded cable,

$$\begin{aligned} V_1 &= 1/\sqrt{L_1C_1} = 1/\sqrt{\mu_1K_1}\times 1/\sqrt{L_0C_0} \quad (2)\\ &= 1/\sqrt{\mu_1K_1}\times 3\times 10^{10} \text{ cm/sec.}\\ &= 1/\sqrt{\mu_1K_1}\times V_0 \end{aligned}$$

where the subscript 1 refers to conditions in the practical, loaded cable, and the subscript 0 refers to conditions in the "ideal" cable of (45).

It may similarly be shown from (31), (47), (35), and (36) that

$$\lambda_1 = 2\pi/\beta_1 = 2\pi/\sqrt{\mu_1K_1}\,\beta_0 = (1/\sqrt{\mu_1K_1})\lambda_0 \quad (3)$$

where the subscripts refer to conditions in the ideal, and the loaded, practical cable, as in (2).

While μ_1 and K_1 will vary slightly with frequency, as do R and G, it is possible to obtain good compensation over a fairly wide band of frequencies.

Pages 121–122, (Fig. 2 and discussion):

This clear figure and the accompanying discussion make it very easy to visualize the formation of standing waves on the line, and contribute greatly to an understanding of this and related phenomena.

Page 122, Paragraph 8 (infinitely long line):

After the excellent discussion of Fig. 2, it is difficult to understand how Mr. Dingley could make the statements contained in this paragraph. These statements apparently arise from a misconception of the true significance of l in (49) and (50). There is some evidence of confusion in paragraph 4 on this page, where "$l=l$" is referred to three different times, yet the context indicates that the meaning is not necessarily the same in each case!

A little consideration will reveal that the l in these equations is the distance from the load, or receiving, end of the transmission line to the point at which it is desired to find the phase and amplitude of E or I.

It cannot be overemphasized that l is in *no way related* to the *ultimate length of the line!*

Reference to Fig. 2, and to the equations, will show that as one moves along the line away from the load, the "reflected" voltage, and current, becomes more and more attenuated, as its vector ($\frac{1}{2}E_2e^{-\alpha l}e^{-j\beta l}e^{j\phi 2}$ in (51)) follows the spiral (Fig. 2) inwards towards zero as an asymptotic value at "infinity." At the same time the "sent"-voltage (and current) vector will follow an outward path along the spiral, becoming "infinite" at "infinity."

The maxima and minima are *in no way* affected by *any* conditions existing in that section of the line which is on the "generator side" of the section under consideration, so long as E_r (and I_r) *is maintained.* The use of the transmission line as an impedance-measuring device is based on this fact.

Page 124, Paragraph 8 (numerical example):

In order for the computations to check, it is necessary for Z_{oc} to be $322-j540$ ohms.

Page 125:

The value of $277e^{j88.7°}$ is $6.3+j277$.

E. N. Dingley, Jr.:[2] I appreciate Mr. Heath's kindness in commenting on my paper. His comments aid materially in clarifying several otherwise obscure points. In apology for my lack of clarity and for the inexcusable arithmetical error on pages 124 and 125, I would like to add that the subject paper was written only for my own edification and use about eight years ago and was dusted off and submitted only in response to the Institute's urgent appeals for material.

Page 120, Paragraphs 8 and 13: α, the "attenuation constant," might better be defined as "the natural logarithm of the ratio of the amplitude of the transmitted voltage (or current), measured at any point on the line, to the amplitude of the transmitted voltage (or current)

[2] United States Navy Radio and Sound Laboratory, San Diego, California.

measured at a second point one unit length away from the first point in the direction of flow of the transmitted energy, in the absence of a reflected wave." The reflected wave alone is attenuated in exactly the same ratio. The vector sum of the transmitted and reflected waves add in phase at certain points of the line to produce a maximum of voltage (or current) and they oppose in phase at other points to produce a minimum of voltage (or current). The resulting sinusoidal voltage (or current) variations, as the point of measurement progresses along the line, are called "standing waves." Due to the presence of standing waves, the direct measurement of α is difficult. It is most easily calculated by means of (58).

On page 124, column 2, lines 11, 15, 26, and 32, the number "560" should, in each instance, be "541," and in line 11, a minus sign should be placed before the j in the exponent of e.

Institute News and Radio Notes

Board of Directors

September 5 Meeting: At the regular meeting of the Board of Directors, which was held on September 5, 1945, the following were present: W. L. Everitt, president; G. W. Bailey, executive secretary; S. L. Bailey, W. L. Barrow, E. F. Carter, W. H. Crew, assistant secretary; Alfred N. Goldsmith, editor; R. F. Guy, R. A. Hackbusch, R. A. Heising, treasurer; F. B. Llewellyn, B. E. Shackelford, D. B. Sinclair, W. O. Swinyard, H. M. Turner, H. A. Wheeler, L. P. Wheeler, and W. C. White.

Building-Fund Campaign Report: Chairman Shackelford reported a Building-Fund total of $601,122, tabulated as follows:

	Actual Production	*As Credited*
Initial Gifts	$506,192.50	$476,732.50
Sections and Individuals	94,929.51	124,389.51
	$601,122.01	$601,122.01

Actual Section subscriptions coming from the members will run very close to the goal of $75,000. The quota for "Initial Gifts" which had been set at $425,000 is now passed and the Building-Fund Committee estimates $625,000 will be received in total.

Constitution and Bylaws

Article IV, Section 2: It was approved to delete Section 2 of Article IV which now reads:

> "Article IV, Section 2—The annual dues shall be payable in advance on the first day of January,"

and to amend it as follows:

> "Article IV, Section 2—When an applicant for membership is elected, the membership period shall be dated as of the first day of the month following election. The member's annual dues period and the period during which he shall receive publications of the Institute shall run concurrently with his membership period. The annual dues shall be payable in advance at the beginning of the annual dues period."

Article IV, Section 3: The recommendation of the Executive Committee that, in accordance with Section 3, Article IV of the Constitution, the Board of Directors waive, until January 1, 1946, all changes in dues of the membership as called for under the recently adopted "Westman Amendment" was unanimously approved

Amended Bylaw Section 46: The following wording was unanimously approved for Bylaw Section 46:

> "Sec. 46—The standing committees, each of which shall normally consist of five or more persons, shall include the following:
>
> Admissions
> Appointments
> Awards
> Board of Editors
> Constitution and Laws
> Education
> Executive Committee of the Board of Directors
> Investments
> Membership
> Nominations
> Papers
> Public Relations
> Sections
> Tellers
> Annual Review
> Antennas
> Circuits
> Electroacoustics
> Vacuum Tubes
> Handbook
> Facsimile
> Frequency Modulation
> Industrial Electronics
> Medical Electronics
> Radio Receivers
> Radio Transmitters
> Radio Wave Propagation and Utilization
> Railroad and Vehicular Communication
> Research
> Standards
> Symbols
> Television
>
> "These committees shall be advisory to the Board of Directors on those matters which are reasonably described by the committee names, except as defined in these Bylaws.
>
> "The terms of appointments of the Admissions, Awards, Board of Editors, Constitution and Laws, Education, Investments, Membership, Nominations, Papers, Public Relations, Sections, and Tellers Committees shall start with the first day of the month following appointment and continue until the date the succeeding terms of appointments take effect. The Board may specifically advance or delay the terminating date of any committee and the starting date of a succeeding committee. The Board shall make appointments to the following committees: Annual Review, Antennas, Circuits, Electroacoustics, Vacuum Tubes, Handbook, Facsimile, Frequency Modulation, Industrial Electronics, Medical Electronics, Radio Receivers, Radio Transmitters, Radio Wave Propagation and Utilization, Railroad and Vehicular Communication, Research, Standards, Symbols, and Television, each year between January first and May first, and the terms of appointments shall be from May first of the year when the appointments are made until April thirtieth of the following year. Additional appointments may be made to fill vacancies or to care for special cases as the need arises, with the term of the appointment expiring April thirtieth."

Admissions Committee Manual: A number of changes in the Admissions Manual were suggested.

Papers Procurement Committee: The tabulation of papers from the survey made by this Committee shows that 164 members are writing papers; 172 plan shortly to do so; and 189 will prepare papers when security regulations permit. Thus, a total of 525 papers can be expected for the PROCEEDINGS.

I.R.E. Representative on RTPB: Dr. W. L. Barrow as nominated as I.R.E. Representative on RTPB, and Dr. D. B. Sinclair was appointed as Alternate I.R.E. Representative.

The Royal Commission on Education: Mr. Hackbusch requested and received permission to mail to The Royal Commission on Education a copy of the reprint of Dr. Everitt's article "The Presentation of Technical Developments Before Professional Societies."

Resumption of Engineering Training: Dr. Everitt called attention to the serious situation caused by the cessation of engineering training during the war and the continuation of drafting of eighteen- to twenty-five-year-old men for the Services, with the resultant interruption in their education. As a result of this situation, Dr. Everitt stated he had called a meeting of the presidents of the leading professional organizations to discuss actions necessary to bring about the resumption of engineering training.

The following resolution was unanimously approved

(*Continued on page 814*)

1946 Winter Technical Meeting

High-lighting the first postwar Winter Technical Meeting of the Institute at the Hotel Astor, January 23 to 26, 1946, four major features are expected to make the session one of the most significant ever held.

As was common at former meetings, Wednesday, the opening day of the convention, will be devoted to I.R.E. business, incuding the Sections Representatives' Meeting and the Sections Representatives' luncheon. The commercial exhibits will open at 6:00 P.M.

Scheduled for Thursday, January 24, are three major features of the Meeting—the Symposium of I.R.E. Technical Committees and technical papers on the latest electronics developments. This year, papers on many vital subjects hitherto barred by military security will be read. Tentative subjects thus far scheduled will include: Broadcasting, Frequency Modulation and Television; Navigational Aids; Communications and Relay Links; Radar; Industrial Electronics; Testing Equipment; new developments in Panoramic Reception; Microwave-Measuring Devices; Broadcast Receivers; Vacuum Tubes; Antennas; and Radio Wave Propagation. As is customary, all papers will have been presented for the first time at this Meeting and none will have been published before in any form. In accordance with last year's successful plan, two technical sessions will be run simultaneously, but wherever possible the papers and sessions will be so arranged that important sessions on the same or related subjects will not conflict. On this same day the Annual Meeting of the Institute will be held. At this time the retiring president will hand the gavel to the incoming president.

One of the outstanding presentations of the Meeting will be the unusually extensive commercial exhibits. They will require all of one floor and part of another in the Hotel Astor. It is expected that 150 or more firms will sponsor this great assembly of exhibits; and so much interest has been shown thus far by radio manufacturers and electronics companies that it seems likely the exhibit space will be completely reserved before the opening day. This timely exhibition will constitute the first radio engineering show of postwar radio equipment and parts. There will be no standard size for any exhibit this year, and firms may select for their exhibits any size they desire within reasonable limits. It is felt that this type of display will benefit the radio engineers and the radio industry as a whole to the greatest extent at this time when the industry is in the midst of reconversion plans.

A third major and enjoyable event of the meeting will be the annual banquet held Thursday evening, January 24, at which a speaker of national prominence will address the members and their guests. In addition, there will be entertainment high lights. At this function, as in former years, the two principal annual awards will be made: the Institute Medal of Honor awarded in recognition of distinguished service in radio communications, and the Morris Liebmann Memorial Prize made "to a member of the Institute who has made public during the recent past an important contribution to radio communications." Announcement will then be made of the election of new Fellows of the Institute, and the retiring president of the Institute, Dr. William L. Everitt, will address the convention.

Following the schedule of past years, the fourth major feature is to be the annual President's Luncheon, held Friday, January 25, in honor of the incoming president.

For out-of-town members it is contemplated that organized inspection trips to points of interest throughout New York city will be arranged. Women's activities, under the direction of Miss Helen M. Stote, are being planned, and will include luncheons and sight-seeing trips.

The Registration desk will function during the entire Meeting. The Exhibits will close promptly at 2 P.M. Saturday, January 26. No organized luncheon is planned for Saturday.

Officials and members of the General Committee for this annual Winter Technical Meeting are: Chairman, Edward J. Content; Vice-Chairman, James E. Shepherd; Secretary, Elizabeth Lehmann; Committeemen, Austin Bailey, George W. Bailey, Stuart L. Bailey, Howard S. Frazier, and William B. Lodge.

Subcommittee chairmen in charge of the various activities are: Arrangements, George Milne; Banquet, C. M. ("Buck") Lewis; Exhibits, H. F. ("Hank") Scarr; Finance, Raymond F. Guy; Hospitality, Philip F. Siling; Papers, Arthur E. Harrison; Printed Program, Dorman D. Israel; Publicity, Will Whitmore; Registration, Harold P. Westman; Section Activities, George B. Hoadley; Special Features, Donald H. Miller; Standing Committee Activities, George T. Royden; Technical Committee Activities, William H. Crew; Women's Committee, Helen M. Stote.

Members and their families are cordially invited to this 1946 Winter Technical Meeting, which is expected to be one of the most pleasant gatherings in the Institute's history, and are urged to make their reservations early. A full-page advertisement appears on page 1A of the advertising section of this issue. At the bottom of this advertisement is a coupon for the convenient use of members in making their hotel reservations.

"The Board of Directors of The Institute of Radio Engineers believes that the immediate resumption of engineering and scientific training is of paramount importance to the economic and military security of the Country, and, to this end, recommends to the President and Congress that exemption from military service should be granted to undergrdauate and graduate students in good standing in approved engineering and scientific curricula. It further recommends that the induction of eighteen-year-old high school graduates accepted for entrance to such curricula should be postponed until an opportunity is given to complete one term, after which the preceding recommendation as to good standing should apply.

"It further urges that release from military service be accelerated for all students whose training in engineering and science has been interrupted."

Executive Committee

September 5 Meeting: The Executive Committee meeting, held on September 5, 1945, was attended by W. L. Everitt, president; G. W. Bailey, executive secretary; S. L. Bailey, W. L. Barrow, E. F. Carter, W. H. Crew, assistant secretary; Alfred N. Goldsmith, editor; and R. A. Heising, treasurer.

Membership: Executive Secretary Bailey reported that, by order of the Board, it was now incumbent upon the Executive Committee to consider applications not approved by the Admissions Committee.

Sixteen applications for transfer to Senior Member grade; three for admission to Senior Member grade; thirty-eight for transfer to Member grade; twenty-seven for admission to Member grade; one hundred and nineteen applications for Associate grade; and thirty-three applications for Student grade were approved and will be found on page 44A of the October, 1945, issue of the PROCEEDINGS.

Constitutional Amendment: President Everitt read the report of the Tellers Committee on the count of ballots on the "Westman Amendment" and commented on the favorable vote of 83.3 per cent for and 16.7 per cent against adoption. It was unanimously approved that the Executive Committee recommend to the Board that, in accordance with Section III, Article 4, of the Constitution, the Board of Dircctors waive until January 1, 1946, all changes in dues of the membership as called for under the recently adopted Westman Amendment.

Connecticut Valley Section: The request of the Connecticut Valley Section for admission to membership in The Connecticut Technical Council was unanimously approved.

Committee Appointment: Dr. Karl Spangenberg was appointed to the Papers Procurement and Papers Committees.

Institute Representatives in Colleges—1945

Alabama Polytechnic Institute: Appointment Later
Alberta, University of: J. W. Porteous
Arkansas, University of: Appointment Later

British Columbia, University of: H. J. MacLeod
Brooklyn, Polytechnical Institute of: G. B. Hoadley

California Institute of Technology: S. S. Mackeown
California, University of: L. J. Black
Carleton College: Appointment Later
Carnegie Institute of Technology: Appointment Later
Case School of Applied Science: P. L. Hoover
Cincinnati, University of: W. C. Osterbrock
Colorado, University of: Appointment Later
Columbia, University of: J. R. Ragazzini
Connecticut, University of: Appointment Later
Cooper Union: J. B. Sherman
Cornell University: True McLean

Detroit, University of: Appointment Later
Drexel Institute of Technology: Appointment Later
Duke University: W. J. Seeley

Florida, University of: P. H. Craig

Georgia School of Technology: M. A. Honnell

Harvard University: E. L. Chaffee

Idaho, University of: H. E. Hattrup
Illinois Institute of Technology: C. S. Roys
Illinois, University of: A. J. Ebel
Iowa, State University of: L. A. Ware

Johns Hopkins University: Ferdinand Hamburger, Jr.

Kansas State College: Karl Martin
Kansas, University of: G. A. Richardson

Lawrence Institute of Technology: H. L. Byerlay
Lehigh University: Appointment Later
Louisiana State University: Appointment Later

Maine, University of: W. J. Creamer, Jr.
Manhattan College: Appointment Later
Maryland, University of: G. L. Davies
Massachusetts Institute of Technology: E. A. Guillemin and W. H. Radford
McGill University: F. S. Howes
Michigan, University of: L. N. Holland
Minnesota, University of: O. A. Becklund

Nebraska, University of: F. W. Norris
Newark College of Engineering: Solomon Fishman
New Mexico, University of: W. F. Hardgrave
New York, College of the City of: Harold Wolf
New York University: Philip Greenstein
North Carolina State College: W. S. Carley
North Dakota, University of: Appointment Later
Northeastern University: G. E. Pihl
Northwestern University: R. E. Beam
Notre Dame, University of: H. E. Ellithorn

Ohio State University: E. C. Jordan
Oklahoma Agriculture and Mechanical College: H. T. Fristoe
Oregon State College: A. L. Albert

Pennsylvania State College: G. L. Crossley
Pennsylvania, University of: C. C. Chambers
Pittsburgh, University of: Appointment Later
Princeton University: J. G. Barry
Purdue University: R. P. Siskind

Queen's University: H. H. Stewart

Rensselaer Polytechnic Institute: H. D. Harris
Rice Institute: Appointment Later
Rose Polytechnic Institute: Appointment Later
Rutgers University: J. L. Potter

Southern Methodist University: Appointment Later
Stanford University: Appointment Later
Stevens Institute of Technology: F. C. Stockwell

Texas, University of: E. W. Hamlin
Toronto, University of: Appointment Later
Tufts College: A. H. Howell

Union College: F. W. Grover
Union States Naval Academy: G. R. Giet
United States Military Academy: P. M. Honnell
Utah, University of: O. C. Haycock

Virginia, University of: L. R. Quarles
Virginia Polytechnic Institute: R. R. Wright

Washington, University of: A. V. Eastman
Washington University: Stanley Van Wambeck
Wayne University: G. W. Carter
Western Ontario, University of: G. A. Woonton
West Virginia University: R. C. Colwell
Wisconsin, University of: Glenn Koehler
Worcester Polytechnic Institute: H. H. Newell

Yale University: H. M. Turner

Technical Committees

May 1, 1945–May 1, 1946

Annual Review

L. E. Whittemore, *Chairman*

R. S. Burnap
W. G. Cady
P. S. Carter
C. C. Chambers
I. S. Coggeshall
W. T. Cooke
D. E. Foster
E. A. Guillemin
R. F. Guy
Keith Henney
I. J. Kaar
A. C. Keller
J. F. Morrison
E. W. Schafer
J. A. Stratton
H. A. Wheeler
C. J. Young
P. D. Zottu

Antennas

P. S. Carter, *Chairman*

Andrew Alford
W. L. Barrow
G. H. Brown
Harry Diamond
W. S. Duttera
Sidney Frankel
R. F. Guy
W. E. Jackson
E. C. Jordan
R. W. P. King,
W. B. Lodge
J. F. Morrison
D. C. Ports
S. A. Schelkunoff
J. C. Schelleng
D. B. Sinclair
George Sinclair
Norman Snyder
L. C. Van Atta

Circuits

E. A. Guillemin, *Chairman*

H. W. Bode
Cledo Brunnetti
C. R. Burrows
F. C. Everett
W. L. Everitt
L. A. Kelley
J. M. Miller
C. A. Nietzert
A. F. Pomeroy
J. B. Russell, Jr.
Stuart W. Seeley
W. N. Tuttle
H. A. Wheeler

Electroacoustics

A. C. Keller, *Chairman*

B. B. Bauer
S. J. Begun
R. P. Glover
F. V. Hunt
V. N. James
R. A. Miller
G. M. Nixon
Benjamin Olney
H. F. Olson
H. H. Scott

Facsimile

C. J. Young, *Chairman*

J. C. Barnes
F. R. Brick, Jr.
Henry Burkhard
J. J. Callahan
A. G. Cooley
R. C. Curtiss
C. N. Gillespie
J. V. L. Hogan
L. D. Prehn
Hugh Ressler
Arthur Rustad
L. G. Stewart
W. E. Stewart
R. J. Wise

Frequency Modulation

C. C. Chambers, *Chairman*

J. E. Brown
M. G. Crosby
C. W. Finnigan
W. F. Goetter
A. C. Goodnow
R. F. Guy
D. L. Jaffe
M. H. Jennings
V. D. Landon
H. B. Marvin
D. B. Smith
J. E. Young

Handbook

H. A. Wheeler, *Chairman*

C. T. Burke
R. S. Burnap
J. D. Crawford
W. H. Crew
R. L. Dietzold
D. G. Fink
Sidney Frankel
Knox McIlwain
Frank Massa
R. D. Rettenmeyer
J. C. Schelleng
F. E. Terman
B. F. Wheeler
J. R. Whinnery
R. M. Wilmotte

Industrial Electronics

P. D. Zottu, *Chairman*
(Appointments Later)

Medical Electronics

(Appointments Later)

Piezoelectric Crystals

W. G. Cady, *Chairman*

C. F. Baldwin
W. L. Bond
J. K. Clapp
Clifford Frondel
E. W. Johnson
W. P. Mason
R. A. Sykes
K. S. Van Dyke
J. M. Wolfskill

Radio Receivers

D. E. Foster, *Chairman*

G. L. Beers
J. E. Brown
W. F. Cotter
L. F. Curtis
W. L. Dunn
H. C. Forbes
C. J. Franks
J. K. Johnson
C. R. Miner
Garrard Mountjoy
H. O. Peterson
H. P. Westman
R. M. Wilmotte

Radio Transmitters

J. F. Morrison, *Chairman*

M. R. Briggs
H. R. Butler
Harry Diamond
F. A. Gunther
W. E. Jackson
W. B. Lodge
J. C. Schelleng
Robert Serrell
D. B. Sinclair
I. R. Weir
J. E. Young

Radio Wave Propagation and Utilization

J. A. Stratton, *Chairman*

S. L. Bailey
C. R. Burrows
T. J. Carroll
D. E. Kerr
H. O. Peterson
J. A. Pierce
S. A. Schelkunoff
R. L. Smith-Rose
H. W. Wells

Railroad and Vehicular Communication

W. T. Cooke, *Chairman*

A. E. Abel
T. G. M. Brown
F. C. Collings
C. N. Kimball, Jr.
D. E. Noble
G. H. Phelps
F. M. Ryan
Winfield Salisbury

Research

F. E. Terman, *Chairman*
(Appointments Later)

Standards

R. F. Guy, *Chairman*

R. S. Burnap
W. G. Cady
P. S. Carter
C. C. Chambers
I. S. Coggeshall
W. T. Cooke
D. E. Foster
Virgil M. Graham
E. A. Guillemin
Keith Henney
L. C. F. Horle
I. J. Kaar
A. C. Keller
J. F. Morrison
Knox McIlwain
E. W. Schafer
H. L. Spencer
J. A. Stratton
H. M. Turner
H. A. Wheeler
L. E. Whittemore
C. J. Young
P. D. Zottu

Symbols

E. W. Schafer, *Chairman*

R. R. Batcher
M. R. Briggs
R. S. Burnap
C. R. Burrows
H. F. Dart
J. H. Dellinger
E. T. Dickey
H. S. Knowles
O. T. Laube
A. F. Pomeroy

Television

I. J. Kaar, *Chairman*

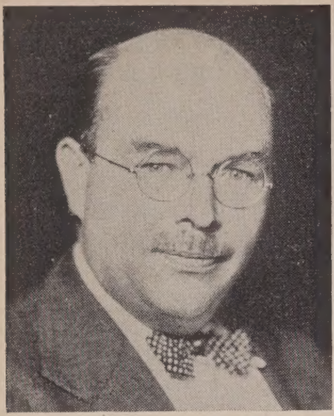
J. E. Brown
K. A. Chittick
D. E. Foster
P. C. Goldmark
T. T. Goldsmith, Jr.
R. N. Harmon
D. L. Jaffe
A. G. Jensen
P. J. Larsen
L. M. Leeds
H. M. Lewis
Jerry Minter
Garrard Mountjoy
D. W. Pugsley
R. E. Shelby
D. B. Sinclair
D. B. Smith

Vacuum Tubes

R. S. Burnap, *Chairman*

E. L. Chaffee
K. C. De Walt
W. G. Dow
R. L. Freeman
A. M. Glover
T. T. Goldsmith, Jr.
J. W. Greer
L. B. Headrick
E. C. Homer
D. R. Hull
F. B. Ingram
D. E. Marshall
J. A. Morton
I. E. Mouromtseff
L. S. Nergaard
G. D. O'Neill
H. J. Reich
A. C. Rockwood
A. L. Samuel
J. R. Steen

I.R.E. People



Lawrence C. F. Horle

Lawrence C. F. Horle

Lawrence C. F. Horle (A'14-M'23-F'25), who has been appointed chief engineer of the Radio Manufacturers Association, engineering department, will be responsible for the management of the department, including the RMA Data Bureau and related activities. Dr. W. R. G. Baker (A'19-F'28) announced Mr. Horle's appointment on August 13, 1945. He stated that the RMA Board of Directors had authorized such personnel and other changes in the engineering department as would be necessary to serve the electronic industry after the war.

Mr. Horle has been associated with the development of the radio industry in various public and private capacities since 1906. Among some of the important positions he has held in the industry are: expert radio aide with the Navy Department's yard at Washington; chief engineer of the de Forest Radio Telephone and Telegraph Company at New York; consultant, Department of Commerce Radio Laboratory, Bureau of Standards, Washington; chief engineer, Federal Telephone and Telegraph Company, New York; and vice-president, Federal Telephone Manufacturing Corporation, Buffalo; president, Radio Club of America; and president, The Institute of Radio Engineers in 1940. He has been employed as a consulting engineer since 1932.

Mr. Horle was graduated from Stevens Institute of Technology. He is a Fellow of the American Institute of Electrical Engineers and the Radio Club of America.

❖

Karl Troeglen

Karl Troeglen, (A'30-M'42-SM'43) on Spetember 1, 1945 joined the KCMO Broadcasting Company of Kansas City, Mo., in the capacity of technical director. Well-known to midwest broadcasters, Mr. Troeglen has been active in commercial radio since 1927. Previous to joining KCMO he was associated with the Western Electric Company, Inc., in the field engineering department. He was a member of the National Association of Broadcasters Engineering Committee during 1941–1942.

Change of Member Address For 1946 Yearbook

If you have made any *changes* in your *address* or *position* since you sent in your Yearbook postcard questionnaire, will you please inform I.R.E. Headquarters promptly. It would be helpful in that case if you would make a statement to the effect that your Yearbook listing should be changed as follows: (Here insert the proper changes.) No Yearbook corrections in address or position can be made after January 1, 1946. Kindly address your communications to

The Institute of Radio Engineers, Inc.
330 West 42nd Street
New York 18, New York

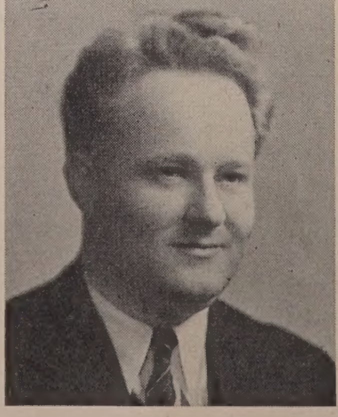
Admiral Stone Honored

The United States Army Distinguished Service Medal was presented on August 9, 1945, to Rear Admiral Ellery W. Stone (A'14-M'16-F'24), USNR, of New York City, Chief Commissioner, Allied Commission, by Vice-Admiral William Glassford, Commander of the United States Navy's Eighth Fleet, at a ceremony in the office of the Mediterranean Theater's Deputy Supreme Allied Commander, General Joseph T. McNarney.

The Citation, approved by the President of the United States, reads as follows:

"Rear Admiral Ellery W. Stone earned the admiration and respect of his Allied associates and members of the Italian government for his outstanding services from September, 1943, to May, 1945, with the Allied Military Mission, later the Allied Control Commission, successively as Director of Communications subcommission, Vice-President, Deputy Chief Commissioner and Chief Commissioner. With the Communications Subcommission, Admiral Stone, through his alertness and grasp of the situation, was able to plan, co-ordinate and execute the restoration of communications in liberated Italy. Later, as Vice-President and

❖



Karl Troeglen

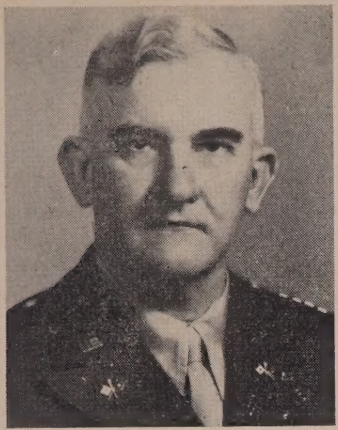
Ellery W. Stone

Deputy Chief Commissioner, he was senior representative of the Allied Control Commission at Salerno, then the seat of the Italian government in liberated territory. There he dealt directly with the Italian government and was responsible for the general enforcement and execution of the surrender terms and for the insurance that the Italian government's conduct would conform to the requirements of an Allied base of operations. As Acting Chief Commissioner and, from November, 1944, as Chief Commissioner, Admiral Stone had full executive responsibility for the activities of the Allied Commission during the period beginning shortly after the fall of Rome and ending with the complete liberation of Italy. With great foresight, he planned military government activities in the north and formed organizations to meet the difficult political problems encountered. At the same time, he pressed steadily and effectively for the assumption of greater responsibility by the Italian government."

Among those present at the ceremony were Field Marshal Sir Harold R. L. G. Alexander, Supreme Allied Commander, and General Joseph T. McNarney, and other high American and British generals and admirals of the Mediterranean command.

At a ceremony at the Palazzo Quirinale in Rome, Italy on August 10, 1945, Rear Admiral Ellery W. Stone, USNR, Chief Commissioner of the Allied Commission, was presented with the Order of Knight of the Grand Cross of St. Maurice and St. Lazarus by the Lieutenant General of the Realm, Crown Prince Umberto. Present at the investiture were Prime Minister Parri, Foreign Minister DeGasperi, and Undersecretary of the Presidency of the Council of Ministers Arpesani. Prime Minister Parri, who as "General Maurizio" of the Italian Partisans of Northern Italy had been in close collaboration with Admiral Stone for many months prior to the German surrender, spoke at length of the assistance given to Italy by Admiral Stone since September, 1943. Founded in the twelfth Century, the Order of Grand Cross of St. Maurice and St. Lazarus is the highest award of knighthood conferred in Italy.

Contributors



Roger B. Colton

Roger B. Colton was born on December 15, 1887, at Jonesborough, North Carolina. He was graduated from Sheffield Scientific School, Yale University, in 1908, and received the degree of Master of Science from the Massachusetts Institute of Technology in 1920. He served details as a special Army student at Massachusetts Institute of Technology and Columbia University, having entered the Army as a Second Lieutenant in 1910.

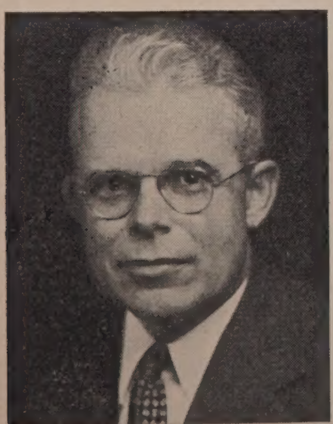
In 1934, Major-General Colton was placed in charge of the research and development division of the Signal Corps Laboratories, later serving as the director of the Signal Corps Laboratories; and, with the advent of the war, as chief of the signal supply service and chief of the engineering and technical service of the Signal Corps.

In 1944, coincident with the division of radio and radar responsibilities between the Signal Corps and the Air Forces, General Colton was transferred to the Army Air Forces, and at present is assistant for electronics to the chief of matériel and services, at headquarters.

❖

Enoch B. Ferrell (A'25–M'29–SM'43) was born in Sedan, Kansas, in 1898. He received the B.A., B.S. in E.E., and M.A. degrees, from the University of Oklahoma in 1920, 1921, and 1924, respectively. Mr. Ferrell taught in the department of mathematics of the university until 1924.

Since 1924 Mr. Ferrell has been a member of the research department of the Bell Telephone Laboratories, where he has been engaged in work on short-wave and ultra-short-wave radio transmitters, and on relays and switches for use in the telephone central-office plant.



Enoch B. Ferrell

❖

D. Gabor

D. Gabor was born at Budapest in 1900. He studied at the Technische Hochschule in Berlin and received his doctor's degree in 1927. In 1925, Dr. Gabor constructed one of the first high-speed cathode-ray oscillographs, and in 1926, the first trigger circuit for the automatic oscillography of transients. From 1927 to 1933, he was associated with the physical laboratory of Siemens and Halske, Berlin, engaged in the development of high-pressure quartz lamps, with the molybdenum-foil seal now in general use.

Dr. Gabor has been employed by the research laboratory of the British Thomson Houston Company, in Rugby, England, since 1934. His work deals with low-pressure discharge devices, high-vacuum electronics, and optical problems. He is a Fellow of the Institute of Physics, and was the 1944 winner of the Duddell Premium of the Institution of Electrical Engineers.

❖

Harold Goldberg (A'38–M'44–SM'44) was born in Milwaukee, Wisconsin, on January 31, 1914. He received the B.S. degree in electrical engineering in 1935, the M.S. degree in 1936, and the Ph.D. degree in 1937 from the University of Wisconson. He served as research Fellow in electrical engineering from 1935 to 1937. After teaching engineering mathematics at the University of Wisconsin during 1937 and 1938, Dr. Goldberg received an appointment as post-doctorate research Fellow in physiology at this University and conducted biophysical research under this Fellowship until June, 1941. He received the Ph.D. degree in physiology from the University of Wisconsin in March, 1941. From 1941 to the end of 1944 he was with the Stromberg-Carlson Company research department in the capacity of senior engineer. Since January, 1945, he has been associated with the research and development section of Bendix Radio Division of the Bendix Aviation Corporation. He is a member of the American Institute of Electrical Engineers and Sigma Xi.

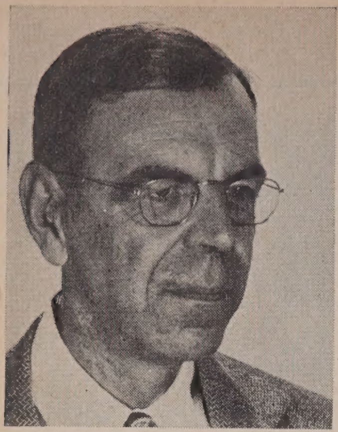
Harold Goldberg

❖

P. M. Honnell (J'27–A'29–M'41–SM'43) was born on January 28, 1908, in Paris, France. He received the B.Sc. in E.E. degree from Texas A. and M. College, in 1930, and subsequently studied at Technische Hochschule, in Vienna; the Conservatoire des Arts et Metiers in Paris; the Massachusetts Institute of Technology, receiving the M.Sc. in E.E. degree; and the California Institute of Technology, from which he received the M.Sc. degree.

From 1926 to 1928 he was a radio operator in the United States Merchant Marine. In 1930 he joined the technical staff of the Bell Telephone Laboratories, remaining with that organization until 1933, when he became a research geophysicist with the Texas Company, a position which he held until 1938.

P. M. Honnell



S. S. Mackeown

❖

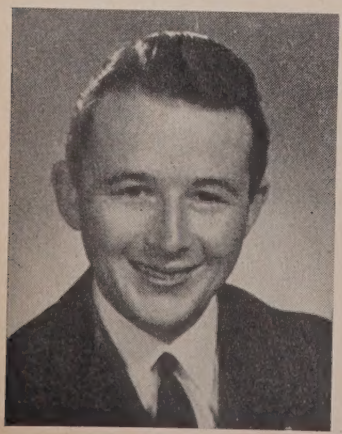
In 1940 Colonel Honnell was appointed assistant professor of electrical engineering at Southern Methodist University. As a reserve officer, he was called to active duty in 1941 as a member of the staff and faculty of the Signal Corps School, at Fort Monmouth, N. J., and in 1942 he was assigned to the faculty of the United States Military Academy, at West Point, N. Y. He holds the rank of Lieutenant Colonel, and at present is in charge of the laboratories and director of the course in electronics of the department of chemistry and electricity at West Point.

❖

William J. Lattin (A'41) was born on July 23, 1910, in Davenport, Iowa. He received the B.S. degree in 1932, and the M.S. degree in 1933, from the Case School of Applied Science.

From 1933 to 1935, Mr. Lattin was an assistant in the electrical engineering department of Columbia University. During 1935 and 1936 he was employed by the Electric Controller and Manufacturing Company, Cleveland, Ohio, as engineer. From 1936 to 1940 he was a radio engineer associated with the Ken-Rad Tube and Lamp Corporation, in Owensboro, Kentucky, now a division of the General Electric Company. From 1940 to 1942, Mr. Lattin served as a radio engineer with the Civil Aeronautics Administration, in Washington, D. C., and in 1942 he returned to the Ken-Rad Corporation, where he has remained to date. He is a member of Sigma Xi, Tau Beta Pi. and Eta Kappa Nu.

❖

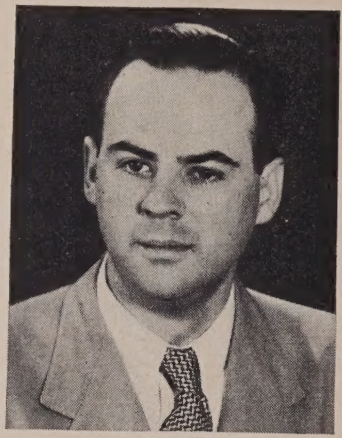


John W. Miles

S. Stuart Mackeown (A'19–M'29–F'40) was born on December 3, 1895, in New York City. He received the B.A. degree from Cornell University in 1917, and the Ph.D. degree in 1923.

From 1918 to 1919 he served as a Second Lieutenant in the radio development section of the Signal Corps. From 1923 to 1926 he was a National Research Fellow at the California Institute of Technology. In 1926 he joined the electrical engineering staff of the California Institute of Technology, where he is now professor of electrical engineering, in charge of work on electronics and communications.

❖



William J. Lattin

❖

John W. Miles was born on December 1, 1920, in Cincinnati, Ohio. He received the B.S. degree in electrical engineering in 1942, the M.S. degree in electrical engineering, the M.S. degree in aeronautical engineering, and the Ph.D. degree in aeronautical engineering, all from the California Institute of Technology.

In the summer of 1942, Dr. Miles was associated with the General Electric research laboratory, and later was a teaching fellow at California Institute of Technology, in Pasadena, California. He was subsequently employed by the Radiation Laboratory at Massachusetts Institute of Technology, and is at present associated with the Lockheed Aircraft Corporation, in Burbank, California.

Dr. Miles is a member of the American Institute of Electrical Engineers, Tau Beta Pi, and Sigma Xi.

❖

Bruce E. Montgomery (S'34–A'38–M'44) was born on July 11, 1913, at Milan, Missouri. He received the A.B. degree from Park College in 1934, the B.S. degree in electrical engineering in 1936, and the degree of Electrical Engineer in 1943, both from Iowa State College.

In 1937, Mr. Montgomery was a student engineer with the Westinghouse Electric and Manufacturing Company. Since 1937 he has been associated with the United Air Lines Transport Corporation, first as an engineer in the communications laboratory and presently as a project engineer in the engineering department. He is a member of Sigma Pi Sigma and Eta Kappa Nu.

Bruce E. Montgomery

❖

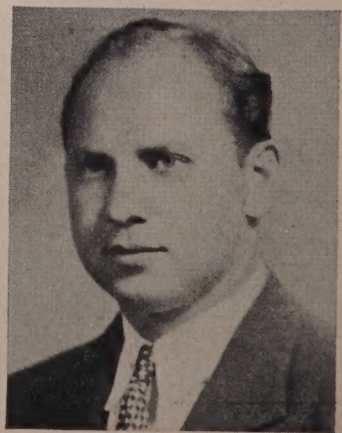
Harry B. Shaper was born in New York City on September 10, 1913. He received the B.S. and M.S. degrees from the School of Technology, College of the City of New York, in 1936, and from 1937 to 1940 he attended the evening sessions of Brooklyn Polytechnic Institute.

Mr. Shaper joined the staff of The Sonotone Corporation in 1936, and until 1940 he was associated with that organization as an engineer working on hearing-aid microphone, phone, and amplifier designs. In 1940 he became associated with The Brush Development Company, developing surface-roughness instruments, microphones, phones, and recording instruments. He is now head of the acoustic and instrument engineering departments of The Brush Development Company.

❖

For photographs and biographical sketches of W. R. Hill, Jr., and Chandler Stewart, Jr., see the January, 1945, issue of the Proceedings; for Howard A. Chinn, see September, 1945; and for Arthur B. Bronwell, see October, 1945.

❖



Harry B. Shaper